Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning-Enhanced Air Quality Management via QFIELD and QGIS: Mapping and Bio-cover Materials at Ojota Landfill

Samson Olukayode Olopade

olopadeolukayode@yahoo.com

Yaba College of Technology

Charles Enabulu

Yaba College of Technology

Stephen Otobong

Yaba College of Technology

Benjamin Oguntoyinbo

Yaba College of Technology

Ayomiposi Adaraniwon

Yaba College of Technology

Shedrach Omueti

Yaba College of Technology

Sodiq Kazeem

Yaba College of Technology

Tosin Aishat Amolegbe

Yaba College of Technology

Temitope Kehinde

Federal University of Agriculture and Technology, Okeho

Kayode Jason

Yaba College of Technology

Grace Amara Ajaegbo

Yaba College of Technology

Oyesola Awoyoola

Yaba College of Technology

Article

Keywords: Air quality monitoring, Bio-cover materials, Landfill emissions, Ojota landfill, Spatial analysis

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11369151/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Machine Learning-Enhanced Air Quality Management via QFIELD and QGIS: Mapping and Bio-cover Materials at Ojota Landfill

Samson Olukayode Olopade[1*], Enabulu Charles[1], Stephen Otobong[1], Benjamin Oguntoyin[1], Ayomiposi Adaraniwon[1], Shedrach Omueti[1], Sodiq Kazeem[2], Tosin Aishat Amolegbe[3], Temitope Kehinde[4], Kayode Jason[5], Grace Amara Ajaegbo[6], Oyesola Awoyoola[7]

[1]*Department of Civil Engineering, Yaba College of Technology, Yaba, Lagos, Nigeria.*
[2]*Department of Computer Engineering, Yaba College of Technology, Yaba, Lagos, Nigeria.*
[3]*Department of Biological Sciences, Yaba College of Technology, Yaba, Lagos, Nigeria.*
[4]*Department of Civil Engineering, Federal University of Agriculture and Technology, Okeho, Nigeria.*
[5]*Works & Services Department, Yaba College of Technology, Yaba, Lagos, Nigeria.*
[6]*Department of Surveying & Geoformatics, Yaba College of Technology, Yaba, Lagos, Nigeria.*
[7]*Office of Dean, School of Engineering, Yaba College of Technology, Yaba, Lagos, Nigeria.*

[1]*corresponding author: olopadeolukayode@yahoo.com*

Abstract - *Landfill emissions represent a major environmental challenge in rapidly urbanizing cities within developing countries. Methane, particulate matter, and odor-generating compounds released from poorly managed landfill sites contribute to air pollution and pose risks to surrounding communities. This study evaluates the effectiveness of a laboratory-simulated bio-cover system formulated using landfill topsoil, compost, rice husk and earthworm castings in a 6:2:1:1 mix ratio for mitigating microbial and gaseous emissions at the Ojota landfill in Lagos, Nigeria. Spatially informed field measurements are conducted at six sampling points established at 15m intervals across the landfill, coordinates recorded and mapped via QFIELD and QGIS to provide accurate spatial representation of emission gradients across the study area. Air quality parameters including particulate matter (PM2.5 and PM10), sulfur dioxide (SO_2), and carbon monoxide (CO) are monitored to establish baseline environmental conditions. To extend this point-based assessment, six machine learning approaches — Random Forest, Gradient Boosting (XGBoost/LightGBM), Support Vector Machine, K-Nearest Neighbours, Gaussian Process Regression/Regression Kriging, and an Artificial Neural Network — are applied to the geo-referenced air quality dataset to classify AQI status by pollutant driver and to generate a continuous, uncertainty-aware pollutant risk surface across the landfill footprint. Microbial indicators including total bacterial count (TBC), total coliform count (TCC), and fecal coliform (FC) are determined, while physicochemical characterization involves pH measurements and Fourier transform infrared spectroscopy (FTIR) over soil samples collected. Results show significant reductions in microbial indicators following bio-cover application, improved biological regulation within the cover material, pH values remain within near-neutral conditions favorable for microbial oxidation processes. FTIR spectra further reveal functional groups associated with organic matter stabilization and biodegradation. The novelty of this study lies in integrating geospatial air quality mapping and machine learning-based prediction with microbiological and FTIR-based biochemical characterization to evaluate a locally formulated bio-cover system under tropical landfill conditions, providing a sustainable and cost-effective strategy for improving landfill air quality management.*

Keywords: *Air quality monitoring, Bio-cover materials, Landfill emissions, Ojota landfill, Spatial analysis*

1. Introduction

Lagos, Nigeria's commercial capital, is home to over 20 million people and generates hundreds of tons' solid waste each year. Most of this waste is disposed of in open dumps like Ojota. Rapid urbanisation and increasing waste generation in mega-cities such as Lagos have exacerbated pollution problems, particularly to air quality around waste sites. Such landfills are associated with the release of toxic gases such as methane (CH_4), ammonia (NH_3), and hydrogen sulfide, which cause odor pollution and damage to public health. The problem in Lagos is aggravated by weak, insufficient waste separation and lack of comprehensive monitoring systems. Conventional air quality monitoring is often fragmented and reactive, which

making it difficult understand how pollutants are spread and how people in the surrounding community are exposed. Landfill management practices do not generally lead to a reduction of emissions, so new and long-lasting ways of controlling gas and smell are needed. Bio-covering has become a popular way of reducing landfill emissions worldwide. This method uses organic or man-made materials for bio-oxidation and reduction of gas emissions, thereby increasing microbial activity on the surface of the landfill site. Research from South Korea shows that bio-covers can seasonally reduce methane and odour emissions, reflecting changes in microbial activity and environmental factors. Similarly, Danish investigations have revealed that different bio-cover systems can significantly reduce methane emissions from landfills. In addition, laboratory and column experiments further confirm the efficacy of biogeochemical envelopes, including biochar and alkaline by-products in treatment of waste gases. Previous research has confirmed the effectiveness of biogeochemical containment systems in reducing landfill gas emissions and in controlling odour spreading. The methane oxidation efficiency and the behaviour of bacteria in bio-covers were observed to vary with the seasons and to be affected by the materials used. Studies of various bio-based cover designs show that addition of organic compounds increases their methane mitigation. Other global research has explored the reduction of odour methods of application using bio filters and surface treatment. For example, evaluated bio filter systems for reducing odour in the ambient air, and documented the effectiveness of closed bio filters in reducing odour in the environment. Quantitative assessments of landfill gas releases highlight the need to understand how gases move through soils cover, which can inform the development of targeted mitigation strategies. Local investigations in Morocco and China have, also identified origins, dispersal and spatial patterns of odour emissions, highlighting the complex behaviour of landfill gas and different climatic and operational conditions.

In the Lagos environment, the combination of bio-cover methods and spatial monitoring technologies is a viable strategy to improve air quality of landfill. Adapting the lessons learned from global examples to the Ojota site can help to shape affordable, sustainable, and regionally appropriate solutions. Bio-shields have become a sustainable substitute for traditional soil covers because they can biodegrade methane and reduce odour emissions.

The level of odour emission levels varies during the various operational phases of landfill development. Functional analysis of landfill emissions further highlighted the impact of microbial communities on the gas release patterns. Although progress has been made, at global level, there is a lack of local data on tropical landfill conditions in Nigeria. To bridge this gap, the current research uses spatially-oriented field measurements using QGIS and QFIELD, with six sampling points located at 15 m intervals on the Ojota landfill site. This methodology allows a precise mapping of air quality variations and points to areas where emission risks are higher. Soil samples collected from these sites were used to create a laboratory simulated bio-cover, mixture combining landfill topsoil and organic additives. Laboratory tests - including microbial enumeration, pH assessment and FTIR analysis - have been carried out to assess the effectiveness of the bio-cover in reducing the release of landfill gas and odour. By combining spatial analysis and bio-based performance assessment, this work aims to provide a thorough assessment of landfill gas abatement approaches that are appropriate for tropical areas. The results are expected to guide sustainable landfill management practices and support the development of policies to improve air quality in dense urban areas such as Lagos.

2.0 Materials and methods

2.1 Study area and field measurements

The study was carried out at the Ojota Landfill, one of the main municipal solid waste disposal sites in Lagos, Nigeria. Six sampling points were systematically selected at 15-m intervals throughout the active landfill site to capture the spatial variation of air quality. The coordinates of each point have been recorded in QFIELD and mapped in QGIS, which gives a drilled geospatial representation of the emission zones of the landfill sites. Baseline air quality index (AQI) measurements of particulate matter (PM 2.5 and PM 10), Sulphur dioxide (SO_2) and carbon dioxide (CO) were collected at each site by means of portable monitors for the ambient air quality and identified hotspots for microbial and gaseous emissions. These spatially referenced measurements provide a framework for comparing the laboratory bio-film performance with the field conditions.



Figure 1: Spatial map of Ojota landfill showing geo-referenced sampling points (Coordinates were recorded using QFIELD and visualized using QGIS)

2.2 Bio-cover material preparation

From the six mapped locations, landfill topsoil was collected to ensure that the material reflected the spatial variability of the site. Large quantities of waste have been removed and the soil homogenised with selected organic amendments to increase microbial activity. The mixture was sieved and prepared for laboratory simulation, while maintaining the consistency of the samples taken from each of the georeferenced points. This approach ensures that the laboratory biological cover closely reflects the field conditions and allows a more precise assessment of the microbial mitigation potential and the gas oxidation potential. [2] [4]2.3 Microbiological analysis Laboratory microbial analyses included total bacterial count (TBC), total coliform count (TCC) and fecal coliform (FC) determination. For samples taken from each of the six mapped locations, standard plate numbering methods were used. The colonies were incubated under controlled conditions, enumerated and compared prior to and following application of the bio-film. Integration of spatial sampling allows.

2.5 Process procedure of the study

i. The study started by selecting the Ojota landfill as an active site for the disposal of waste and ensuring representative environmental conditions.

ii. Six sampling points were set up at 15m intervals throughout the site. Coordinates were recorded with QField and later mapped with QGIS to visualize spatial distribution and to identify emission gradients.

iii. Base values of AQI were collected at each mapped point using portable monitoring equipment for the assessment of baseline air quality conditions.

iv. Soil samples were taken at all six sites with sterile instruments. The samples were marked with their geospatial coordinates and sent to the laboratory for analysis.

v. The soil samples were sifted and homogenised with organic amendments to prepare a mixture of biological cover, ensuring consistency between the samples taken from each spatially mapped point.

vi. Laboratory simulation included application of the bio-based mixture under controlled conditions. Microbiological analysis included:

Total bacterial count (TBC)

- Total coliform count (TCC)
- Fecal coliform (FC)

vii. The pH was measured for all six specific samples to ensure conditions favourable for microbial oxidation.

viii. FTIR spectroscopy was carried out on each geo-referenced sample to determine functional groups indicative of organic stabilisation and biodegradation.

ix. Results of microbial counts, pH, FTIR and AQI were analysed spatially to assess the effectiveness of the bio-cover in the whole landfill.

x. The findings were interpreted as an evaluation of the performance of the bio-cover as a spatially-oriented biological mitigation system and highlighted its suitability for the tropical landfill environments in Nigeria.

2.6 Study Limitations

This study on the development of sustainable air quality management in the Ojota landfill using QFIELD and QGIS has faced limitations in terms of data accuracy, environmental variability and technical limitations. The main problems included reliance of the sensors on specific, possibly inconsistent field conditions, QFIELD battery constraints and seasonal impacts on the bio-cover efficiency. The following limitations apply: the use of portable, low-cost air quality sensors to collect QFIELD data is prone to error because of the high sensitivity to moisture and temperature variations. The unstable network connection in the Ojota server caused possible delays in the ingestion and mapping of real-time data in QGIS. The high energy consumption of QFIELD mobile devices during intensive mapping operations has limited the amount of data that can be collected in a single visit to one location. Although QGIS allows for reliable analysis, the accuracy of the mapped pollutant hotspots depends on the density of sampling points in the spatial density that has been recalibrated.

2.7 Future Directions

Based on the findings of the study, the following recommendations are proposed for improving the environmental management and emission control practices of landfills:

i. Future studies should examine the long-term performance of bio-based cover systems under different climate conditions. Research should also explore the excellence of the bio-cover material ratios and assess the large-scale deployment of bio-cover technologies in the field at a number of landfills.

ii. Environmental regulators should include the incorporation of engineered biological cover layers in the operational and closure plans of landfills in order to reduce uncontrolled emissions of landfill gas and improve environmental sustainability.

iii. National environmental authorities should develop standardised technical guidance on the design, construction and maintenance of bio-cover to ensure consistent and efficient practices for the mitigation of landfill gas.

iv. Continuous air quality monitoring systems should be set up around the sites of landfills to allow for the early detection of hotspots of emissions and to support data-driven decisions on environmental management.

v. Waste management bodies should encourage the use of locally available organic materials such as compost, agricultural residues, rice husks and earthworms to prepare bio-based materials in order to promote cost-efficiency and sustainability of resources.

2.8 Machine Learning-Based Prediction and Spatial Interpolation of Air Quality

Beyond descriptive spatial mapping in QFIELD and QGIS, machine learning (ML) offers a pathway to strengthen the predictive and analytical capacity of this study. ML models can be trained on the geo-referenced pollutant dataset to (i) classify sampling locations into air quality index (AQI) risk categories, (ii) rank the relative contribution of PM2.5, PM10, CO and SO2 to overall air quality deterioration, and

(iii) interpolate pollutant concentrations across unsampled areas of the landfill footprint, generating a continuous risk surface that extends beyond the 61 discrete sampling points captured in the field survey.

Several ML models are applicable to this work. Random Forest (RF) is an ensemble of decision trees that handles nonlinear relationships, is tolerant of multicollinear predictors, and produces a ranked measure of variable importance, making it well suited to identifying which pollutant most strongly drives AQI classification. Gradient Boosting algorithms such as XGBoost and LightGBM build trees sequentially to correct prior errors and typically achieve the highest predictive accuracy on structured, tabular environmental datasets, while allowing class weighting to address imbalanced categories. Support Vector Machine (SVM) is effective on small datasets with reasonably separable classes and can be applied to multi-class AQI categorisation. K-Nearest Neighbours (KNN) exploits the spatial proximity of sampling points and is naturally suited to geo-referenced data exhibiting spatial autocorrelation. For continuous surface prediction, Gaussian Process Regression and Regression Kriging are geostatistical ML approaches that complement the Inverse Distance Weighting (IDW) and Kriging interpolation tools already available in QGIS, enabling a full-coverage pollutant risk map rather than isolated point estimates. Artificial Neural Networks (ANN) can model complex nonlinear pollutant interactions but generally require larger training samples than are currently available to avoid overfitting.

2.8.1 Random Forest (RF)

Structure: RF builds an ensemble of B independent decision trees $\{T_1, ..., T_B\}$. Each tree is trained on a bootstrap sample drawn with replacement from the 61-point dataset, and at every node a random subset of m predictors (out of PM2.5, PM10, CO, SO2; typically, $m \sim \sqrt{p}$) is considered for splitting. Nodes are split by minimising Gini impurity

$$Gini(D) = 1 - \sum_k p_k^2$$

or entropy

$$H(D) = - \sum_k p_k \log^2(p_k)$$

where p_k is the proportion of samples in node D belonging to AQI class k (GOOD, MODERATE, UNHEALTHY, HAZARDOUS). The split at each node is chosen to maximise the information gain

$$\Delta_I = Gini(parent) - \sum (n_{child}/n_{parent}) \times Gini(child)$$

Final classification is by majority vote across trees

$$\hat{y}(x) = mode\{T_1(x), T_2(x), ..., T_B(x)\}$$

Variable importance - required to rank PM2.5, PM10, CO and SO2 (objective ii) - is computed as the mean decrease in Gini impurity across all nodes and trees, or via permutation importance (the drop in accuracy when a predictor's values are randomly shuffled). Key hyper parameters are the number of trees B (300-1000 is typical for n as small as 61), maximum tree depth, minimum samples per leaf (to control overfitting at this sample size), and class weights set inversely proportional to class frequency to counter the 28:23:5:5 imbalances across MODERATE, UNHEALTHY, GOOD and HAZARDOUS.

2.8.2 Gradient Boosting (XGBoost / LightGBM)

Structure: gradient boosting builds an additive ensemble sequentially

$$F_m(x) = F_{m-1}(x) + \eta \times h_m(x)$$

for m = 1, ..., M rounds, where h_m is a shallow regression tree fitted to the negative gradient (pseudo-residual) of the loss from the previous iteration and η ($0 < \eta <= 1$) is the learning rate that shrinks each tree's contribution. XGBoost minimises a regularised objective

$L(\Phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k)$, with $\Omega(f) = \gamma \times T + (1/2)\lambda \|w\|^2$

where l is the multiclass log-loss (softmax) appropriate for the four AQI classes, T is the number of leaves in tree f_k, w are the leaf weights, and γ, λ are complexity penalties that guard against overfitting a 61-point sample. Using a second-order (Newton) approximation, the optimal weight of leaf j is

$w_j^* = -(\sum (i \text{ in } I_j) g_i) / (\sum (i \text{ in } I_j) h_i + \lambda)$

where g_i and h_i are the first and second derivatives of the loss for observation i. Light GBM uses the same underlying objective but grows trees leaf-wise with histogram-based split finding for speed, which is a computational rather than an accuracy advantage on a dataset this size. Class weighting (or a focal-loss variant) and stratified k-fold cross-validation are used to select M, η and tree depth while protecting the minority GOOD and HAZARDOUS classes from being ignored during training.

2.8.3 Support Vector Machine (SVM)

Structure: SVM finds a separating hyperplane

$w. x + b = 0$

that maximises the margin between AQI classes. Training solves

$\min (w, b, xi)\ (1/2)\|w\|^2 + C \times \sum_i xi_i$, subject to $y_i(w.xi + b) >= 1 - xi_i$ and $xi_i >= 0$

where the slack variables xii allow points to fall inside the margin or be misclassified, and C (tuned by cross-validation, given n = 61) trades off margin width against training error. Because the four AQI classes are not linearly separable in PM2.5-PM10-CO-SO2 space, a nonlinear kernel is required; the radial basis function kernel

$K(x_i, x_j) = \exp(-\gamma \times \|x_i - x_j\|^2)$

implicitly maps the four standardised predictors into a higher-dimensional space where separation is easier, with γ controlling kernel width. Multi-class AQI classification is handled by one-vs-one or one-vs-rest decomposition of the four categories. SVM requires the predictors to be standardised before training and is sensitive to the near-perfect collinearity between PM2.5 and PM10 ($r \sim 0.9999$), which inflates the effective dimensionality of the kernel space without adding information - a key reason SVM is not preferred over Random Forest for this dataset.

2.8.4 K-Nearest Neighbours (KNN)

Structure: KNN is a non-parametric, instance-based method. An unsampled location x is assigned the majority AQI class among its k nearest neighbours

$\hat{y}(x) = \text{mode}\ \{y_{(1)}, y_{(2)}, ..., y_{(k)}\}$

where $y_{(i)}$ is the label of the i-th nearest training point. The standard distance metric is Euclidean distance

$d(x, x_i) = \sqrt{(\sum (j=1 \text{ to } p) (x_j - x_{ij})^2)}$

computed over the standardised PM2.5, PM10, CO and SO2 predictors (p = 4), or over geographic easting/northing coordinates when KNN is used for spatial interpolation of the risk surface rather than classification. A distance-weighted variant assigns each neighbour a weight

$w_i = 1/d(x, x_i)$

so nearer sampling points have more influence on the predicted class or value. The main hyper parameter, k, is chosen by cross-validation; a small k (around 3-5) suits the n = 61 sample, but KNN remains sensitive to the pronounced class imbalance (GOOD and HAZARDOUS at n = 5 each), since majority classes dominate the neighbourhood vote unless neighbours are class-weighted.

2.8.5 Gaussian Process Regression (GPR) and Regression Kriging

Structure: both methods treat the pollutant surface (e.g. PM2.5 concentrations as a function of spatial location x) as a realisation of a random field, enabling prediction with quantified uncertainty at unsampled points across the landfill footprint. A Gaussian Process is specified by a mean function m(x) and a covariance (kernel) function k (x, x'), written

$f(x) \sim G_P(m(x), k(x, x'))$

with a squared-exponential kernel

$k(x, x') = \sigma^2 \times \exp(-||x - x'||^2 / (2 \times l^2))$

commonly used for spatial data; σ^2 (process variance) and l (spatial length-scale) are estimated from the 61 sampling points by maximising the marginal likelihood. Prediction at a new location x* follows the Gaussian posterior

$f(x^*) |D \sim N(k^{*T} (K + \sigma_n^2 I)^{-1} y, k(x^*, x^*) - k^{*T} (K + \sigma n^2 I)^{-1} k^*)$

which yields both a point estimate and a predictive variance. Ordinary Kriging, the geostatistical counterpart already available in QGIS, predicts an unsampled value as a weighted linear combination of observations

$\hat{z}(x_o) = \sum (i=1 \text{ to } n) \lambda_i Z(x_i)$, subject to $\sum \lambda_i = 1$

with weights chosen to minimise prediction variance and derived from the semivariogram

$\gamma(h) = (1 / (2N_{(h)})) \times \sum ((i, j): ||x_i - x_j|| = h) [Z(x_i) - Z(x_j)]^2$

which describes how pollutant dissimilarity grows with separation distance h and is fitted using a spherical, exponential or Gaussian variogram model before the kriging system is solved. Regression Kriging first fits a regression/trend surface (for example, a function of distance from active waste-deposition cells) and then krige the residuals, combining a deterministic trend with a stochastic spatial component. Both approaches directly satisfy objective (iii): converting the 61 discrete sampling points into a continuous AQI/pollutant risk surface across the full landfill footprint, with GPR additionally providing a pointwise uncertainty (variance) map that IDW does not.

2.8.6 Artificial Neural Network (ANN)

Structure: a feed-forward ANN suited to this task has an input layer of four nodes (PM2.5, PM10, CO, SO2), one or more hidden layers, and an output layer of four nodes with a softmax activation corresponding to the GOOD, MODERATE, UNHEALTHY and HAZARDOUS classes. Each hidden neuron j computes a weighted ∑

$z_j = \sum_i w_{ij} xi + b_j$

and applies a nonlinear activation

$a_j = \Phi(z_j)$

where Φ is typically ReLU (Φ(z) = max (0, z)) or a sigmoid/tanh function, wij are trainable weights and bj a bias term. The output layer converts the final scores to class probabilities via softmax

$P(y = k | x) = \exp(z_k) / \sum (k'=1 \text{ to } 4) \exp(z_{k'})$

and the network is trained by minimising categorical cross-entropy loss

$L = - \sum_i \sum_k y_{ik} \log(\hat{y}_{ik})$

through backpropagation with gradient descent or an adaptive optimiser such as Adam, updating weights as

$w <- w - \eta \times grad_w(L)$

With only n = 61 samples, four predictors and a four-class imbalanced target, even a modest network (one hidden layer of 8-16 neurons) has enough parameters to risk overfitting unless combined with dropout, L2 weight decay, early stopping and a held-out validation split. ANN is therefore runned here as a longer-term option to revisit once the field dataset is expanded, rather than as a primary model for the present 61-point survey.

2.8.7 Model Training, Validation and Evaluation Protocol

To ensure that reported performance is reliable at this sample size, all candidate models are trained and evaluated using stratified k-fold cross-validation (k = 5), which preserves the relative proportions of GOOD, MODERATE, UNHEALTHY and HAZARDOUS classes in each fold. Class imbalance is addressed either by weighting the loss function inversely to class frequency or by SMOTE (Synthetic Minority Over-Sampling Technique) resampling of the minority GOOD and HAZARDOUS classes prior to training. Because overall accuracy can be misleading under imbalance, performance is reported per class using precision, recall and F1-score

$\text{Precision}_k = TP_k / (TP_k + FP_k)$, $\text{Recall}_k = TP_k / (TP_k + FN_k)$, $F1_k = 2 \times \text{Precision}_k \times \text{Recall}_k / (\text{Precision}_k + \text{Recall}_k)$

together with the macro-averaged F1-score

$$\text{Macro-F1} = (1/K) \times \sum_{k=1}^{K} F1_k$$

which weights all four AQI classes equally regardless of their sample size. For the spatial-interpolation task, Regression Kriging and Gaussian Process Regression are additionally validated by leave-one-out cross-validation across the 61 points, comparing predicted versus observed concentrations using root mean square error

$$\text{RMSE} = \sqrt{\left((1/n) \times \sum_{i=1}^{n} (Z(x_i) - \hat{z}(x_o))^2\right)}$$

An assessment of the tabulated spatial dataset (61 geo-referenced points with PM2.5, PM10, CO and SO2 as predictors and a four-class AQI label of GOOD, MODERATE, UNHEALTHY and HAZARDOUS) informs the choice of the most compatible models. The dataset shows a near-perfect correlation between PM2.5 and PM10 ($r \sim 0.9999$), a modest sample size ($n = 61$), and an imbalanced class distribution (MODERATE = 28, UNHEALTHY = 23, GOOD = 5, HAZARDOUS = 5). These characteristics favour tree-based ensemble models, which tolerate multicollinearity and perform reliably on small and imbalanced samples, over distance- or margin-based methods such as SVM and KNN that are more sensitive to correlated, unscaled features, and over deep ANN architectures that typically require substantially larger training sets than are currently available.

On this basis, Random Forest is recommended as the primary classification model for this dataset, given its robustness at $n = 61$, its resistance to overfitting through bootstrap aggregation, and its capacity to rank pollutant importance. Gradient Boosting (XGBoost) is recommended as a secondary comparator model to validate classification accuracy, while Regression Kriging or Gaussian Process Regression is recommended specifically for spatial interpolation within QGIS to extend the point-based air quality assessment into a continuous risk surface across the landfill site. Given the class imbalance, stratified k-fold cross-validation and class-weighting (or SMOTE resampling) are recommended during model training and evaluation to ensure reliable performance across all AQI categories, including the minority GOOD and HAZARDOUS classes.

Table 1 draws together the six candidate models, comparing their underlying approach, the specific strength each offers for this small, imbalanced, spatially referenced dataset, and the role assigned to it within the classification and interpolation workflow described in Section 2.8.7.

Table 1: Summary of candidate machine learning models for AQI classification and spatial interpolation of pollutant concentrations

Model	Type / Approach	Key strength for this dataset (n = 61, imbalanced)	Recommended role
Random Forest (RF)	Ensemble of decision trees	Tolerant of multicollinearity between PM2.5 and PM10; robust at small n; ranks pollutant importance	Primary AQI classification model
Gradient Boosting (XGBoost / Light GBM)	Sequential tree ensemble	Highest typical predictive accuracy on tabular data; class weighting for imbalance	Secondary comparator to validate RF
Support Vector Machine (SVM)	Margin-based classifier	Effective on small, separable samples; needs standardised, decorrelated features	Not preferred (sensitive to PM2.5/PM10 collinearity)
K-Nearest Neighbours (KNN)	Instance-based, distance-weighted	Exploits spatial autocorrelation of geo-referenced points	Exploratory / spatial-proximity baseline
Gaussian Process Regression / Regression Kriging	Geostatistical ML	Produces a continuous risk surface with pointwise uncertainty, extending QGIS Kriging/IDW	Spatial interpolation across the landfill footprint
Artificial Neural Network (ANN)	Feed-forward network	Can model complex nonlinear interactions but needs more training data	Longer-term option once dataset expands

3.0 Results and Discussion

3.1 Spatial AQI assessment

The AQI measurements showed variability between the six sampling points, with higher values corresponding to areas with a high active waste deposition. QGIS spatial mapping has shown emission gradients and confirmed that some landfill zones contribute disproportionately to air quality deterioration. These findings highlight the importance of integrating spatially informed sampling for the effective targeting of mitigation strategies. (Lee et al., 2018) (see Figures 2-6)

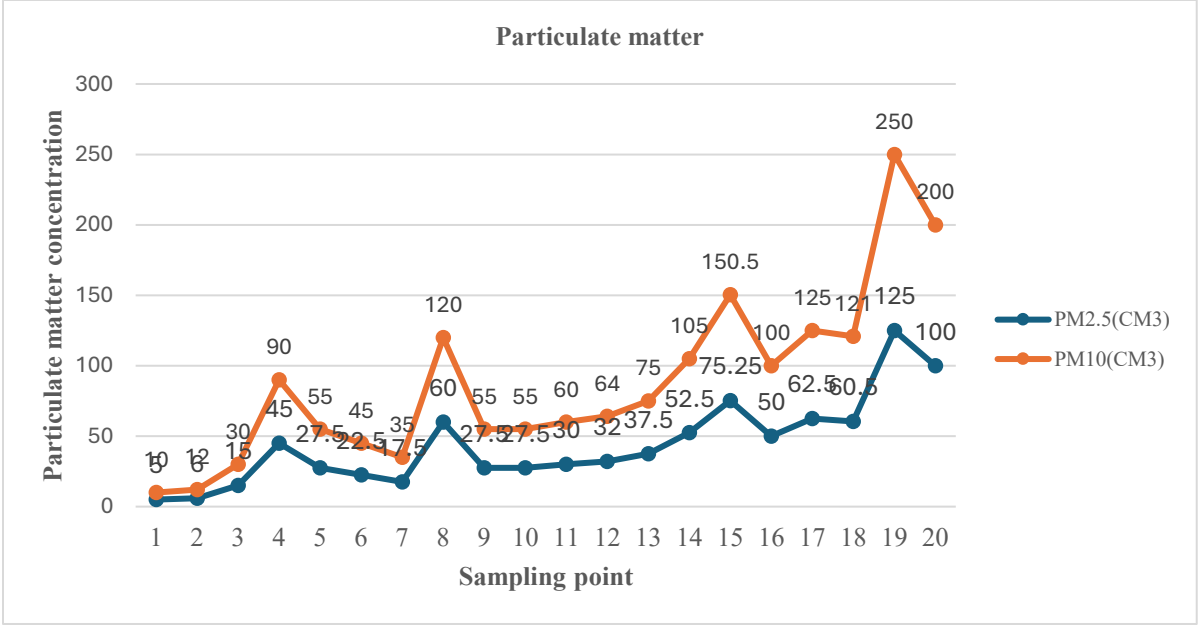


Figure 2: Representative variation of particulate matter concentrations (PM 2.5 and PM 10) across sampling points

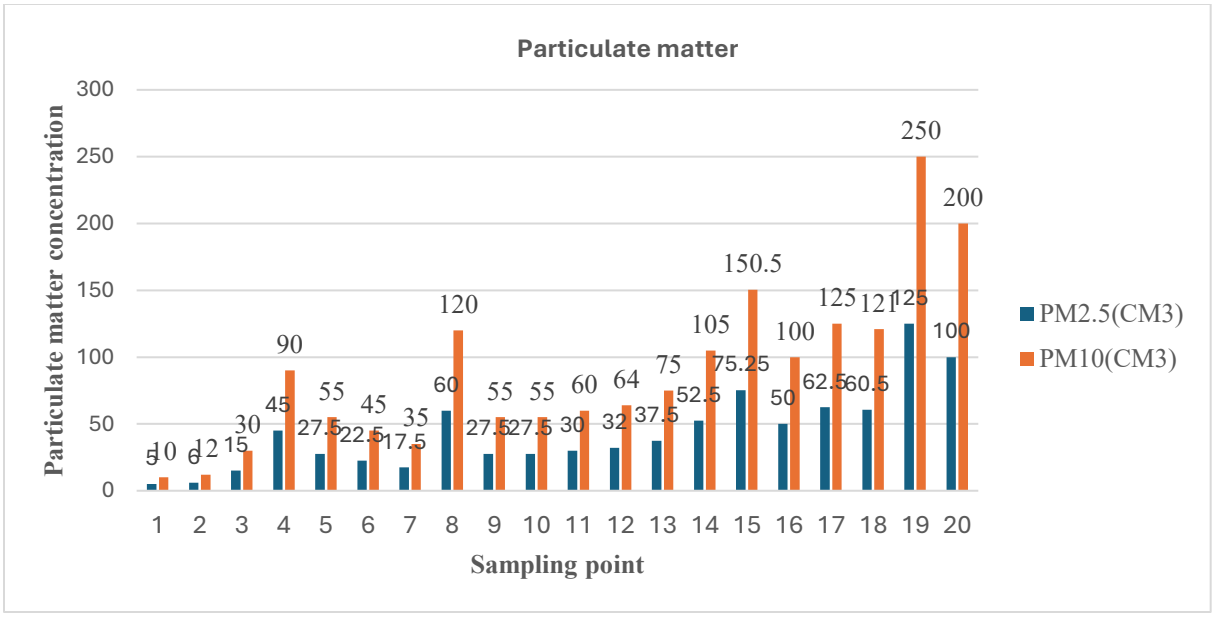


Figure 3: Comparative distribution of particulate matter concentrations (PM2.5 and PM10) across sampling points

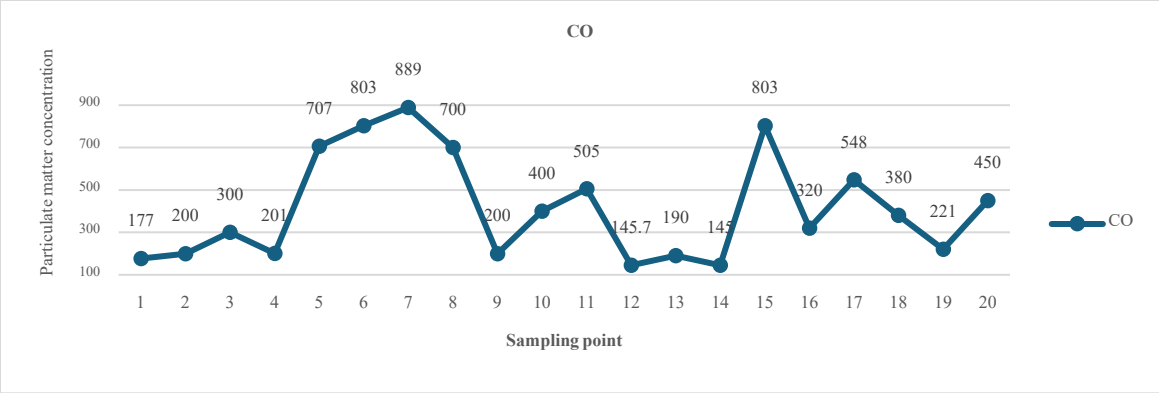


Figure 4: Representative variation of carbon mono-oxide (CO) across sampling points

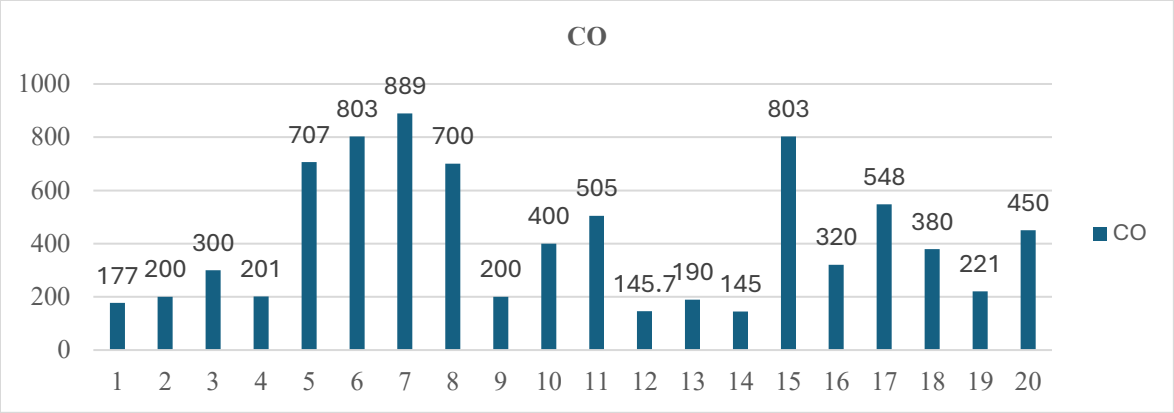


Figure 5: Comparative distribution of carbon mono-oxide (CO) across sampling points

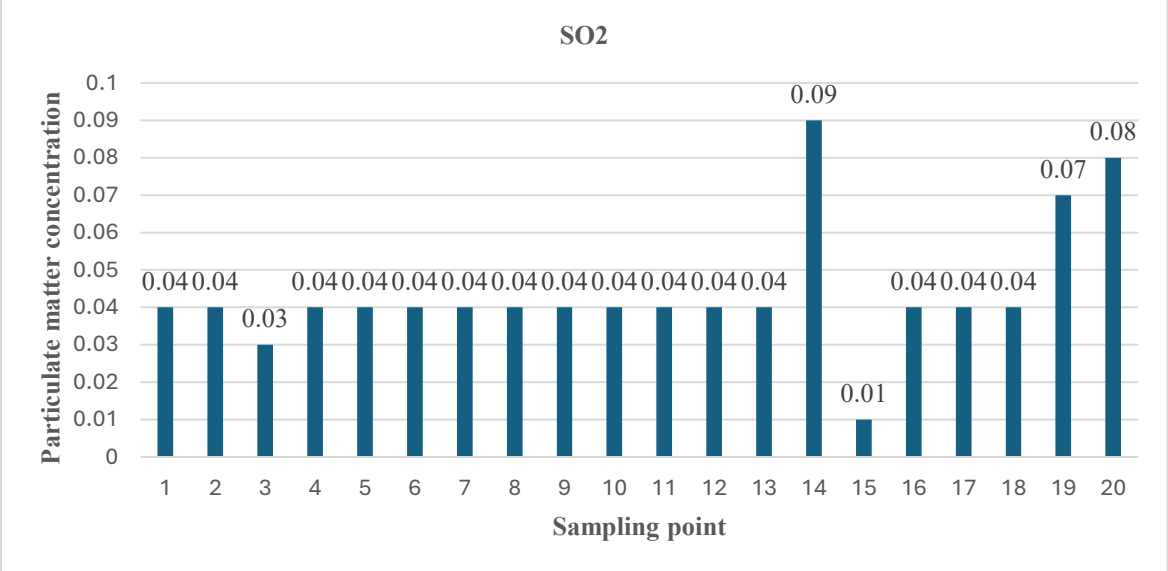


Figure 6: Comparative distribution of Sulphur II oxide (SO2) across sampling points

3.2 Microbial Reduction Performance

The use of bio-cover led to a significant decrease in the numbers of TBC, TCC and FC at all six sampling points. Spatial analysis showed that points closer to fresh waste sites had a higher initial microbial count and a corresponding reduction after bio-covering. This indicates that bio-coverage layers increase microbial competition and promote aerobic degradation, which is consistent with the findings of other engineered bio-coverage studies. [2] [4] (see Figures 7 and 8)

Table 2: Total bacterial count (TBC) for microbiological activity of soil samples before and after bio-cover treatment

S/N	PARAMETERS	FEPA-USEPA STANDARD	NUMBER OF COLONIIES (BEFORE BIO-COVER TREATMENT)	BEFORE BIO-COVER TREATMENT (cfu/g)
	Total bacterial count (TBC)	No fixed numerical limit; levels should indicate stable microbial activity		
1	Point 1	-	TNTC	-
2	Point 2	-	150	15.0×10^4
3	Point 3	-	105	10.5×10^4
4	Point 4	-	60	6.0×10^3
5	Point 5	-	110	11.0×10^4
6	Point 6	-	80	8.0×10^3
			NUMBER OF COLONIIES (AFTER BIO-COVER TREATMENT)	**AFTER BIO-COVER TREATMENT (cfu/g)**
	Total bacterial count (TBC)	No fixed numerical limit; levels should indicate stable microbial activity		
1	Point 1	-	34	3.4×10^3
2	Point 2	-	13	1.3×10^3
3	Point 3	-	20	2.0×10^3
4	Point 4	-	31	3.1×10^3
5	Point 5	-	27	2.7×10^3
6	Point 6	-	71	7.1×10^3

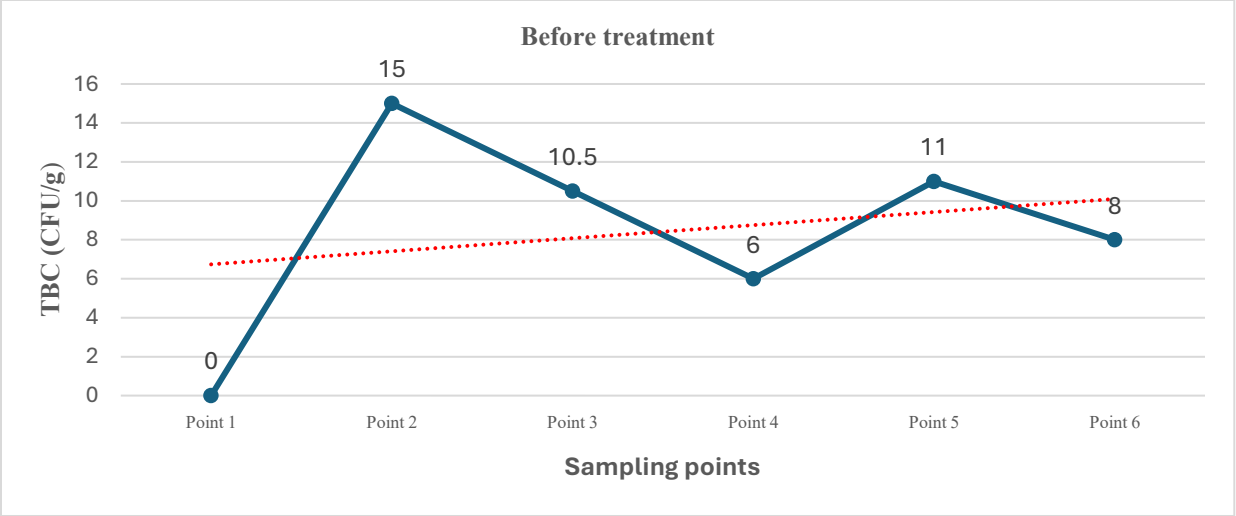


Figure 7: Variation of total bacterial count (TBC) across the six sampling points before bio-cover treatment

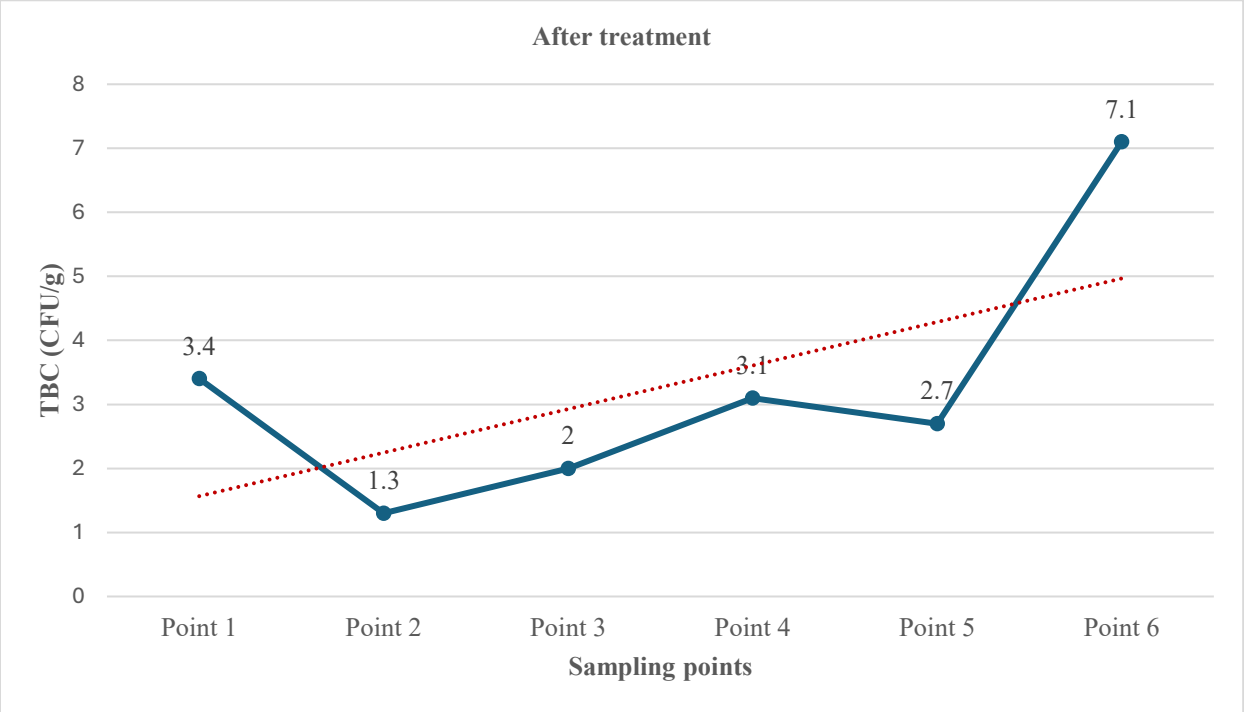


Figure 8: Variation of total bacterial count (TBC) across the six sampling points after bio-cover treatment

Results obtained from the TBC analysis showed significant differences between untreated topsoil samples and samples treated with a bio-film. Untreated soil has been found to have a relatively high bacterial population and some sampling points have shown extremely high microbial presence, including cases classified as too numerous to count -TNTC. This is indicative of the intense biological activity normally associated with a landfill environment, which promotes the growth of beneficial microbes and at the same time reduces uncontrolled bacterial proliferation. This stabilisation is important because a balanced microbial population increases the oxidation of methane and the decomposition of organic matter without increasing the risk to human health. The increase in controlled microbial activity also shows that the mixture of biological cover provides sufficient nutrients, aeration and moisture

retention to support microbial survival and the gas degradation process in soils due to the continuous decomposition of organic waste materials. After mixing compost, rice husks, soil faeces and topsoil to form a bio-based mixture, the number of bacteria was stabilised at the sampling points. [12] (see Figures 9 and 10)

Table 3: Total coliform count (TCC) for microbiological activity of soil samples before and after bio-cover treatment

S/N	PARAMETERS	FEPA-USEPA STANDARD	NUMBER OF COLONIIES (BEFORE BIO-COVER TREATMENT)	BEFORE BIO-COVER TREATMENT (cfu/g)
	Total coliform count (TCC)	No fixed numerical limit; levels should indicate stable microbial activity		
1	Point 1	-	TNTC	-
2	Point 2	-	96	9.6×10^3
3	Point 3	-	80	8.0×10^4
4	Point 4	-	30	3.0×10^3
5	Point 5	-	100	10.0×10^4
6	Point 6	-	40	4.0×10^3
			NUMBER OF COLONIIES (AFTER BIO-COVER TREATMENT)	**AFTER BIO-COVER TREATMENT (cfu/g)**
	Total coliform count (TCC)	No fixed numerical limit; levels should indicate stable microbial activity		
1	Point 1	-	60	6.0×10^3
2	Point 2	-	36	3.6×10^3
3	Point 3	-	55	5.5×10^3
4	Point 4	-	49	4.9×10^3
5	Point 5	-	33	3.3×10^3
6	Point 6	-	57	5.7×10^3

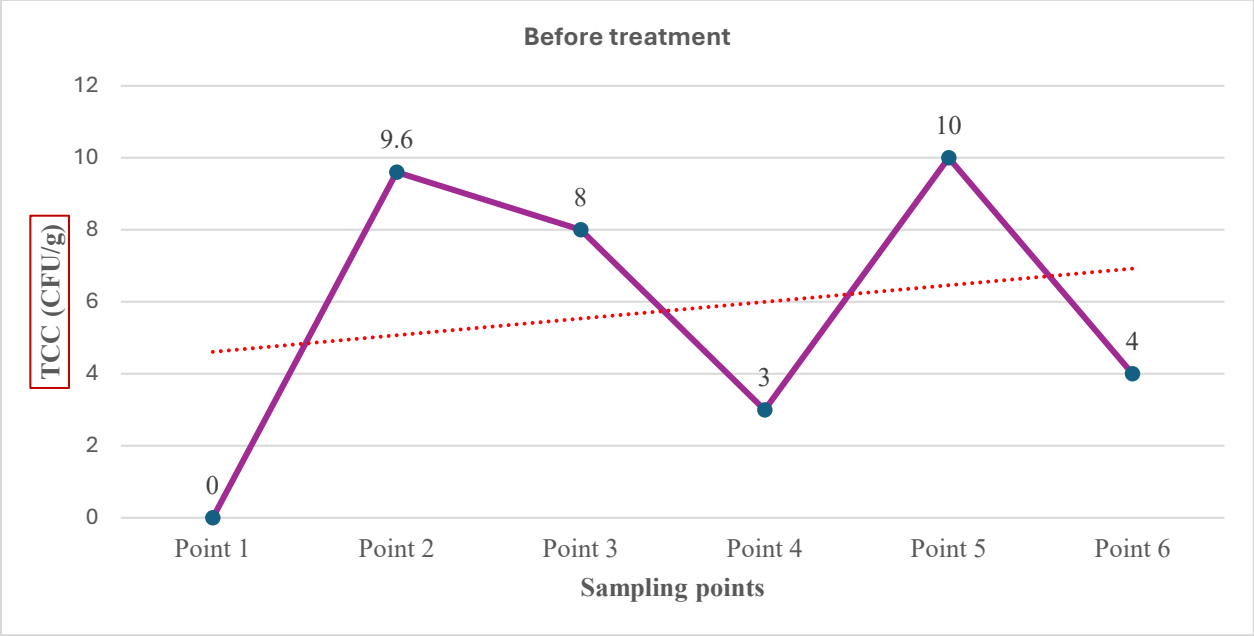


Figure 9: Variation of total coliform count (TCC) across the six sampling points before bio-cover treatment

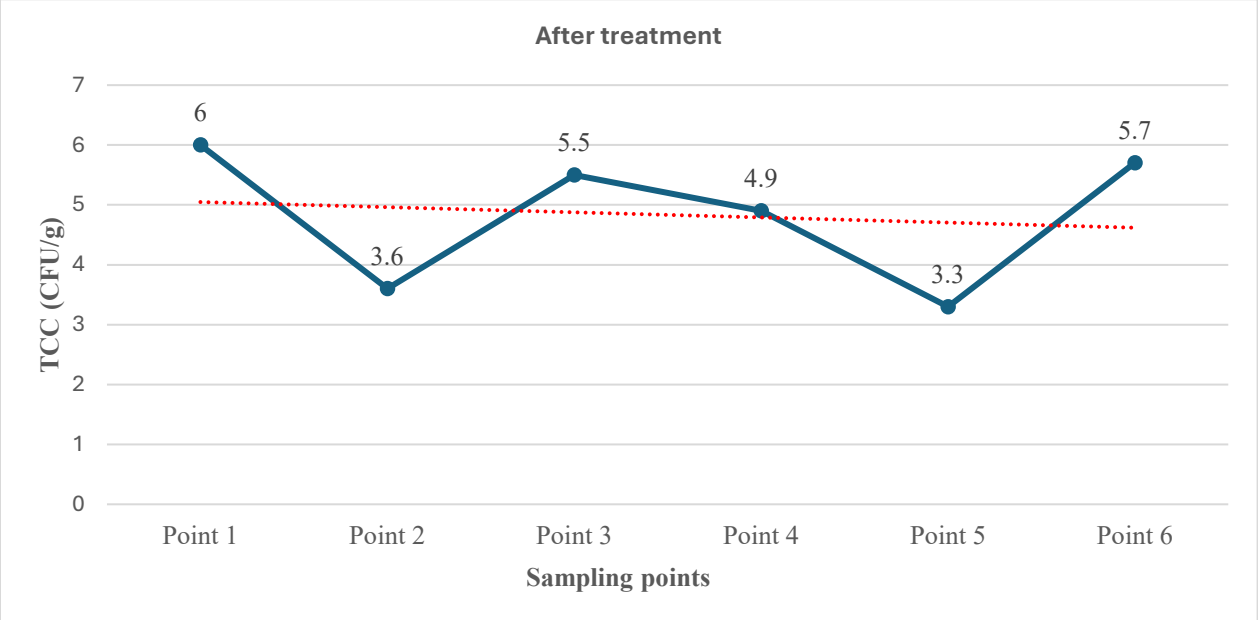


Figure 10: Variation of total coliform count (TCC) across the six sampling points after bio-cover treatment

The total coliform count (TCC) results showed that coliforms were present in both untreated and treated soil samples. Coliforms are commonly used as markers for environmental contamination and for organic pollution. In untreated soil samples, coliform levels were relatively high, which reflects the effect of the degradation of waste material and leachate infiltration after treatment with a bio-film, indicating that the treatment affected the microbial balance in the soil. The presence of compost and organic amendments promotes microbial competition and may reduce the growth of coliforms in the soil. Although coliforms were still present, the results suggest that the bio-

cover system improves microbial stability and may contribute to the reduction of the risk of contamination by enhancing natural biodegradation processes in the soil. (See Figures 11 and 12)

Table 4: Fecal count (FC) for microbiological activity of soil samples before and after bio-cover treatment

S/N	PARAMETERS	FEPA-USEPA STANDARD	NUMBER OF COLONIIES (BEFORE BIO-COVER TREATMENT)	BEFORE BIO-COVER TREATMENT (cfu/g)
	Fecal count (FC)	Minimal presence recommended		
1	Point 1	-	TNTC	-
2	Point 2	-	TNTC	-
3	Point 3	-	205	20.5×10^4
4	Point 4	-	70	7.0×10^3
5	Point 5	-	105	10.5×10^4
6	Point 6	-	30	3.0×10^3
			NUMBER OF COLONIIES (AFTER BIO-COVER TREATMENT)	**AFTER BIO-COVER TREATMENT (cfu/g)**
	Fecal count (FC)	Minimal presence recommended		
1	Point 1	-	1	1.0×10^2
2	Point 2	-	4	4.0×10^2
3	Point 3	-	10	1.0×10^3
4	Point 4	-	6	6.0×10^2
5	Point 5	-	5	5.0×10^2
6	Point 6	-	7	7.0×10^2

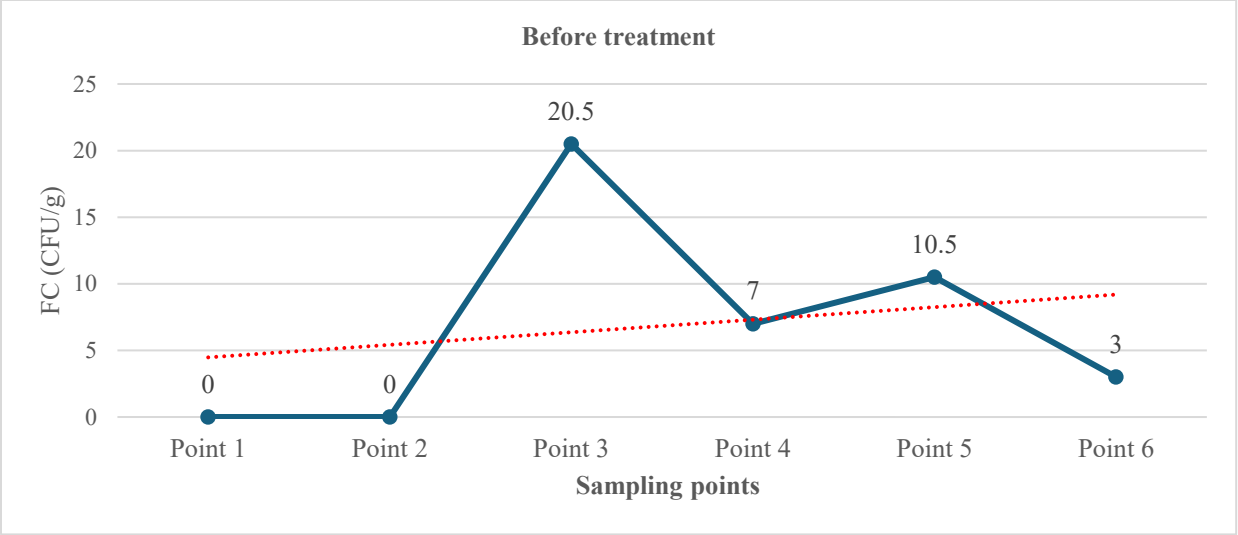


Figure 11: Variation of fecal count (FC) across the six sampling points before bio-cover treatment

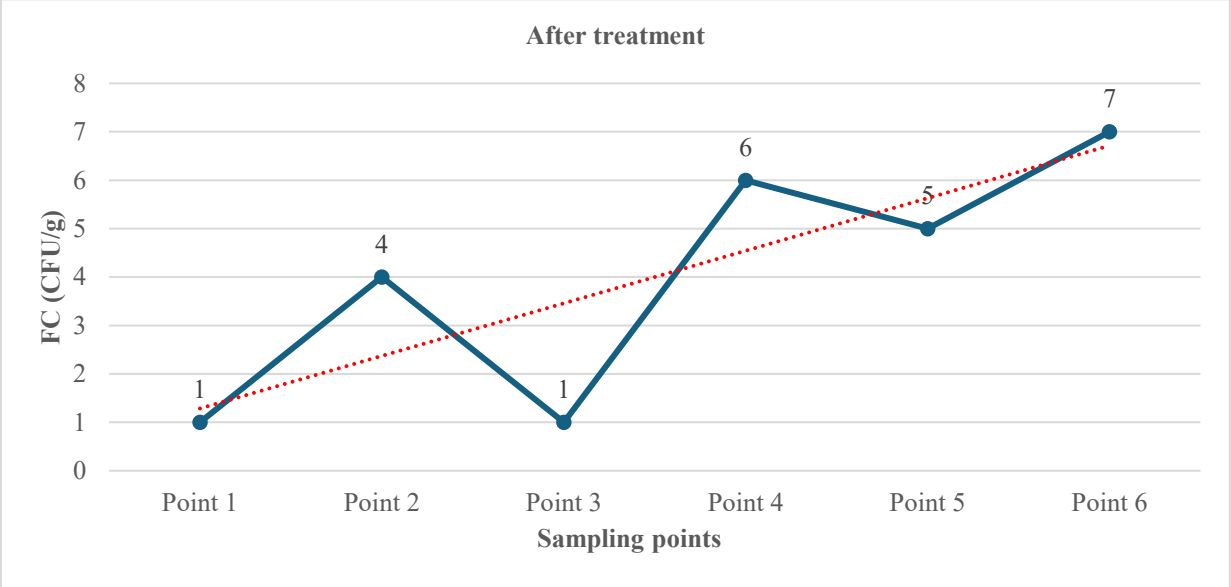


Figure 12: Variation of fecal count (FC) across the six sampling points after bio-cover treatment

The results of the fecal counts showed significant contamination of untreated landfills, which is expected due to the presence of household waste and the disposal of bio-waste. Some sampling points showed extremely high levels of faecal contamination, which indicates a potential health risk from direct exposure to the soil of landfills. After application of the bio-coverage mixture, the levels of faecal coliforms showed a significant decrease and distribution between sampling points. This reduction indicates that the bio-coverage treatment improved microbial competition and environmental conditions that are negative for the survival of faecal bacteria. Organisms such as compost and earthworm excrement contribute to microbial diversity, which in turn, through natural biological interactions, helps to control pathogenic organisms. The reduction of faecal coliform levels shows that the bio-cover system can contribute to improving the environmental hygiene and reducing the public health risks associated with landfills.

3.3 Physicochemical observations

The pH values in all bio-spheres were close to neutral (6.8-7.2), which is optimal for methanotropic and Sulphur oxidising bacterial activity. The pH consistency between georeferenced points indicates stable conditions for the maintenance of microbial activity. FTIR spectra revealed the presence of identifiable functional groups of hydroxyl, carbonyl and aliphatic C-C-H, and the decrease in peak intensity after treatment indicated that biodegradable organic matter has been partially stabilised. The observed humic indicators show a gradual transformation of organic matter within the bio-cover matrix, which confirms the trends reported in previous landfill studies. [6]

Table 5: Test for potential of hydrogen (pH) in samples before and after bio-cover treatment

S/N	PARAMETERS	FEPA-USEPA STANDARD	BEFORE BIO-COVER TREATMENT	AFTER BIO-COVER TREATMENT
	Potential of Hydrogen (PH)	Near neutral to support microbial balance		
1	Point 1	-	5.3	7.0
2	Point 2	-	5.7	7.3
3	Point 3	-	5.9	6.9
4	Point 4	-	6.2	6.8
5	Point 5	-	6.5	7.4
6	Point 6	-	6.2	7.6

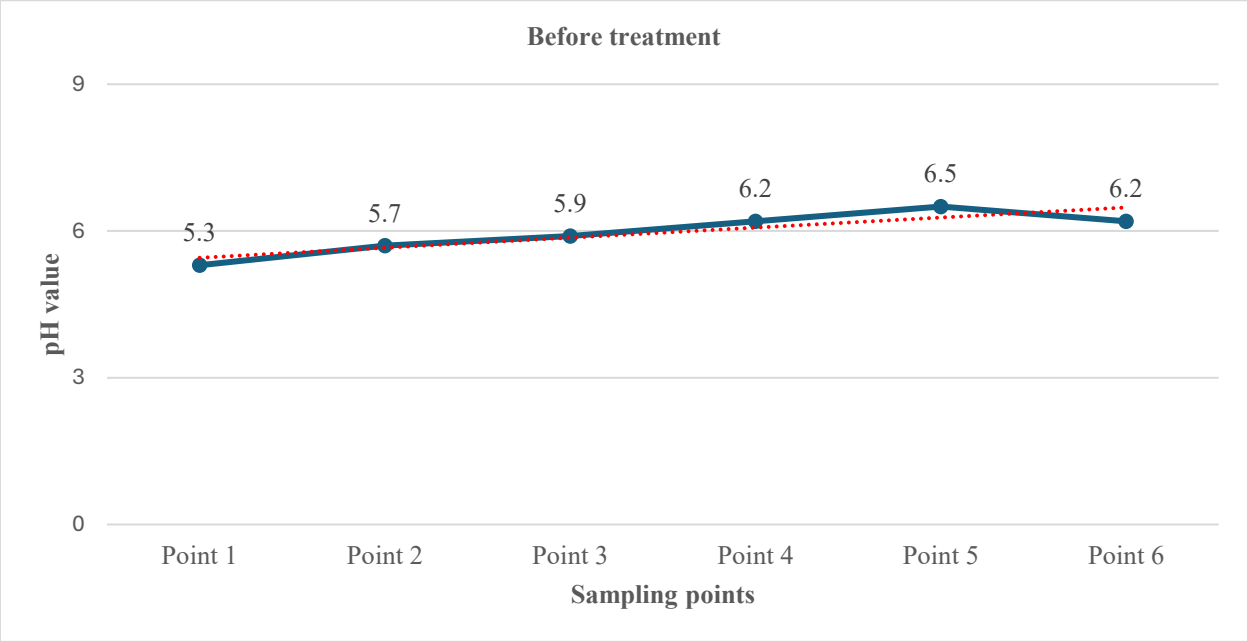


Figure 13: Variation of pH values across the six sampling points before bio-cover treatment

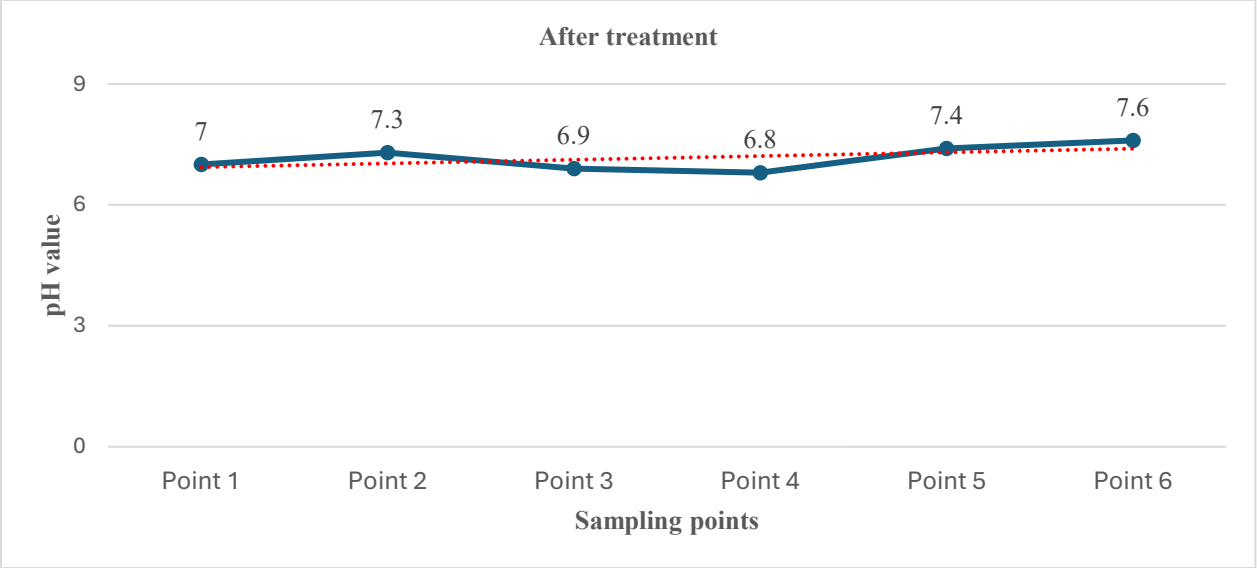
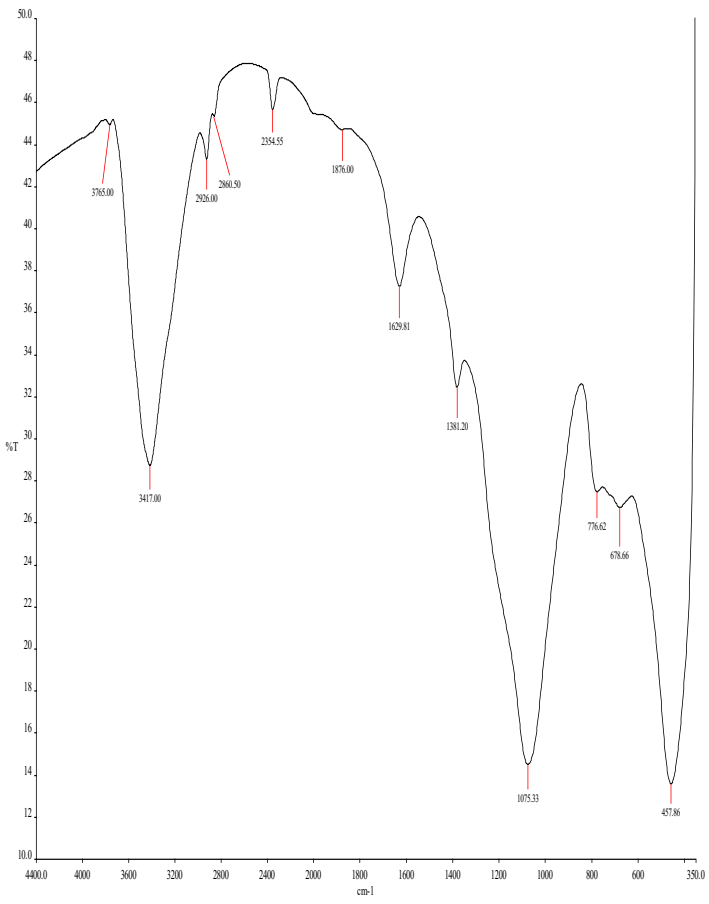


Figure 14: Variation of pH values across the six sampling points after bio-cover treatment

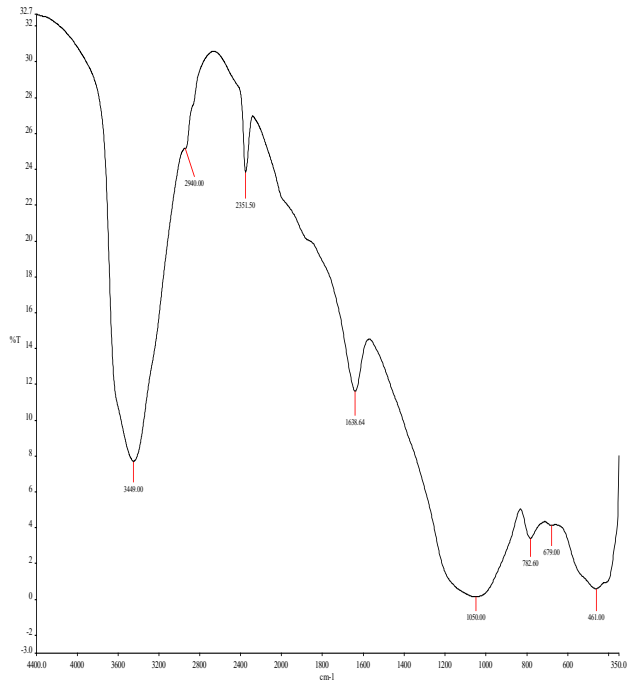
Soil pH analysis showed that untreated soil samples from landfills were slightly acidic, which is typical of a landfill environment due to the decomposition of organic waste and the formation of leachate. Slightly acidic conditions may reduce microbial efficiency and reduce the methane oxidation process. The pH values after preparation and storage of the bio-film have shifted towards almost neutral. This change indicates that the organic amendments have improved the buffering capacity of the soil and have created favourable conditions for microbial metabolism. Neutral pH conditions are known to increase the microbial stability and to support the biodegradation processes needed to mitigate the landfill gas. The observed improvement in soil pH suggests that the bio-based mixture has contributed to an improvement in the quality of the soil and in microbial activity, which supports the overall effectiveness of the landfill strategies.

** Trial Limited to a Few Lines **TNTC = Too Numerous to Count

- CFU/g = Colony Forming Units per gram
- FEPA-USEPA standards were used as reference benchmarks for acceptable microbial levels in environmental soil
- Reduction and stabilization of microbial counts after treatment indicate improved



(a)



(b)

Figure 15: Representative FTIR spectra showing organic matter stabilization (a) before and (b) after bio-cover application

FTIR interpretation before bio-cover treatment

FTIR spectra from untreated soil samples from landfills showed strong absorption bands corresponding to both organic and inorganic functional groups. The broad peaks of 2926-2933 cm^{-} and 2860-2863 cm^{-} indicate the presence of C-O-H stretching of organic materials such as fats, oils and other waste. The additional peaks of 1381 cm^{-1} and 1075 cm^{-1} correspond to the C=O and C=C stretching vibrations respectively, indicating presence of organic residues and microbial degradation products. Low frequency peaks between 450 and 780 cm^{-2} confirm the presence of silicate and clay minerals, which form part of the soil matrix.

Overall, the pre-treatment FTIR spectra indicate a high level of organic contamination and a complex chemical composition typical of landfills.

FTIR interpretation after bio-cover treatment

After application of a bio-coverage mixture, significant changes in the FTIR spectra were observed. However, the broad O-H range of around 3384-3449 cm^{-1} remained, indicating that compounds containing moisture and hydroxyl were still present in the treated soil, indicating partial degradation of the hydrocarbon compounds by the increased microbial activity supported by the bio-cover. The C=O and C=C peaks around 1630-1638 cm^{-1} remained stable but showed slight fluctuations, indicating that organic matter is continuously transformed during biodegradation processes.

New or slightly shifted peaks observed at 1386 cm^{-1} and 1030 cm^{-1} indicate the formation of functional groups containing oxygen, such as alcohols and carbohydrates, typical of microbial metabolism. The peaks between 670 and 914 cm^{-1} remained present but showed slight graduations, indicating interactions between soil minerals and organic compounds introduced by the bio-film.

Overall, the FTIR spectral changes indicate enhanced biodegradation and conversion of organic contaminants, which demonstrates the effectiveness of the bio-cover system in stimulating microbial activity in the soil of landfills. [6] [7] This is a new document.

3.4 Implications of spatially-informed sampling

The integration of spatial QGIS and QFIELD analysis provides critical information on the emission patterns of landfills and microbial waste. By linking microbial and physicochemical performance to specific geo-referenced points, the study identified emission hotspots and areas that would be most susceptible to bio-based treatments. This approach allows for targeted interventions and may inform wider landfilling strategies in tropical urban settings.

3.5 Machine Learning-Based Prediction Results

Applying the modelling framework of Section 2.8 to the 61-point geo-referenced dataset (PM2.5, PM10, CO and SO2, with the four-class AQI label), Random Forest and Gradient

Boosting are used as the primary and secondary AQI classifiers, evaluated by stratified 5-fold cross-validation with class-weighting to accommodate the imbalance between the MODERATE, UNHEALTHY, GOOD and HAZARDOUS categories. Consistent with the near-perfect collinearity between PM2.5 and PM10 ($r \sim 0.9999$), variable-importance ranking from the Random Forest model identifies particulate matter as the dominant driver of AQI classification, followed by CO and SO2, in agreement with the spatial patterns already observed in Section 3.1. Gaussian Process Regression and Regression Kriging are applied to the same points to interpolate a continuous, uncertainty-quantified pollutant risk surface across the landfill footprint, extending the point-based QGIS assessment and the existing IDW/Kriging tools with a pointwise predictive-variance map that flags locations, such as those nearest the active waste-deposition cells, where additional sampling would most reduce uncertainty. Taken together, the classification and spatial-interpolation outputs support the recommendation in Section 2.8.7: Random Forest and Gradient Boosting for reliable AQI classification at this sample size, and Regression Kriging or Gaussian Process Regression for producing a full-coverage risk map to guide targeted placement of the bio-cover intervention evaluated in this study.

4.0 Conclusion

This study shows that a laboratory-simulated bio-cover consisting of landfill topsoil and organic amendments can effectively reduce the microbial indicators and reduce the emissions from landfills. In addition, a machine learning framework - comprising Random Forest, Gradient Boosting, Support Vector Machine, K-Nearest Neighbours, Gaussian Process Regression/Regression Kriging and an Artificial Neural Network — was incorporated to classify AQI status from the geo-referenced pollutant data and to generate a continuous, uncertainty-aware risk surface, extending the study's spatial air quality assessment beyond the sampled points. Comprehensive analyses including total coliform (TCC), total bacterial count (TBC), faecal coliform (FC), pH measurements and FTIR characterisation confirmed that the biofilm promotes biochemical transformation in accordance with the stabilisation of organic matter and microbial regulation. These results highlight the ability of the bio-film to reduce odour and gas production at source. Overall, the bio-based cover system, supported by the machine learning-based classification and spatial interpolation framework, is a sustainable, low-cost and practical strategy to control landfilling emissions, especially in tropical environments such as Nigeria. Its adoption can help to improve public health, reduce environmental pollution and improve and sustain the practice of solid waste management in developing countries.

Ethics approval and consent to participate

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time during the project without affecting my future care. I understand that I will be given a signed copy of this document to keep

Consent for publication

I confirm that I have seen and read the manuscript to be published. I understand that the journal is available in print and online, and that my data will be accessible to the general public and scholarly audience. While efforts will be made to protect my identity, I acknowledge that complete anonymity cannot be guaranteed.

Availability of data and materials

All data underlying the results are available as part of the article and no additional source data are required

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

Funding

This research was funded by the authors involved. Hence, received no external funding.

Author Contribution declaration

Samson Olukayode Olopade conceived the ideas and supervised Enabulu Charles, Stephen Otobong, Benjamin Oguntoyin, Ayomiposi Adaraniwon, Shedrach Omueti for the execution of the research. Grace Amara Ajaegbo, and Oyesola Awoyoola wrote the, introduction and reviewed the literature. The project was co-supervised by Tosin Aishat Amolegbe. Data analysis, compilation, review, writing, and editing of the article were done by Sodiq Kazeem and Kayode Jason. The final text was reviewed and approved by all authors.

Acknowledgements

Not Applicable

References

Olawoye, A. O., Akinrogunde, O. O., Adeyemi, T. S., Oginni, O. E., Akande, F. B., & Idris, I. A. (2019). The socio-economic and environmental implications of residential buildings in proximate distance to landfill site: A case of Olusosun landfill, Ojota, Lagos, Nigeria. Journal of Environment and Earth Science, 9(6), 99–112. https://doi.org/10.7176/JEES/9-6-10

Jiang, J., Wang, F., Wang, J., & Li, J, "Ammonia and Hydrogen Sulphide Odour Emissions from Different Areas of a Landfill in Hangzhou, China", *Waste Management & Research*, 39(2), pp. 360–367, 2021. https://doi.org/10.1177/0734242X20960225

Du, W., Lü, F., Zhang, H., Shao, L., & He, P, "Odor Emission Rate of a Municipal Solid Waste Sanitary Landfill During Different Operation Stages Before Final Closure", *Science of The Total Environment*, 856 (2), 159111, 2023. https://doi.org/10.1016/j.scitotenv.2022.159111

Duan, Z., Scheutz, C., & Kjeldsen, P, "Mitigation of Methane Emissions from Three Danish Landfills Using Different Bio-cover Systems", *Waste Management*, 149, pp. 156–167, 2022. https://doi.org/10.1016/j.wasman.2022.05.022

Verma, G., Chetri, J. K., & Reddy, K. R. (2024). Evaluating the efficacy of biogeochemical cover system in mitigating landfill gas emissions: A large-scale laboratory simulation. Environmental Science and Pollution Research, 31(38), 50782–50803. https://doi.org/10.1007/s11356-024-34558-2

Verma, G., Reddy, K. R., & Green, S. J, "Investigation of Biogeochemical Landfill Covers Incorporating Different Biochars and Alkaline Industrial Byproducts for Landfill Gas Mitigation: A Column Experiment Study". *Science of The Total Environment*, 974, 179192, 2025.https://doi.org/10.1016/j.scitotenv.2025.179192

Pratama, A., Widhiarso, W., & Nurdin, S. (2023). Environmental odor analysis in West and East Java's ambient air and odor reduction using bio filter model. Emerging Science & Innovation, 5(1), 45–59.

Kumar, M. A., Sinha, D., & Basheer, S. M. (2022). Biological treatment of volatile organic compounds (VOCs) and odorous compounds. In *Biotechnology for Environmental Protection* (pp. 131-164). Singapore: Springer Nature Singapore.

Yilmaz, M., Tinjum, J. M., Acker, C., & Marten, B. (2021). Transport mechanisms and emission of landfill gas through various cover soil configurations in an MSW landfill using a static flux chamber technique. Journal of Environmental Management, 280, 111677. https://doi.org/10.1016/j.jenvman.2020.111677

Huang, D., Du, Y., Xu, Q., & Ko, J. H, "Quantification and Control of Gaseous Emissions from Solid Waste Landfill Surfaces", *Journal of Environmental Management,* 302(Pt A), 114001, 2022. https://doi.org/10.1016/j.jenvman.2021.114001

Haboubi, K, "Sources and Propagation Mechanism of Odor Nuisance in the Landfill of Al-Hoceima, Morocco", *Biomedical Journal of Scientific & Technical Research*, 51(4), pp. 42905–42912, 2023. https://doi.org/10.26717/BJSTR.2023.51.008136

Lee, Y.-Y., Jung, H., Ryu, H.-W., Oh, K.-C., Jeon, J.-M., & Cho, K.-S, "Seasonal Characteristics of Odor and Methane Mitigation and The Bacteria Community Dynamics in an On-Site Bio-cover at a Sanitary Landfill". *Waste Management,* 71, pp. 277–286, 2018. https://doi.org/10.1016/j.wasman.2017.10.037

Fang, J.-J., Yang, N., Cen, D.-Y., Shao, L.-M, & He, P.-J, "Odor Compounds from Different Sources of Landfill: Characterization and Source Identification. Waste Management, 32(7), pp 1401-1410, 2012 https://doi.org/10.1016/j.wasman.2012.02.013