



Estimation of Decision Boundaries for Critical Zone Classification in a Polymetallic Tailings Dam Using Machine Learning



Eduardo Manuel Noriega-Vidal^{*}, Jackson Wilder Narvaez-Valdivia[†], Marden Anderson Huancas-Morey[†],
Diego Antonio Hernandez-Puyo[†], Wilberto Effio-Quezada[†]

Department of Mining Engineering, Universidad Privada del Norte, 13011 Trujillo, Peru

* Correspondence: Eduardo Manuel Noriega-Vidal (eduardo.noriega@upn.edu.pe)

Received: 01-20-2026

Revised: 03-04-2026

Accepted: 03-19-2026

Citation: E. M. Noriega-Vidal, J. W. Narvaez-Valdivia, M. A. Huancas-Morey, D. A. Hernandez-Puyo, and W. Effio-Quezada, "Estimation of decision boundaries for critical zone classification in a polymetallic tailings dam using machine learning," *Int. J. Comput. Methods Exp. Meas.*, vol. 14, no. 1, pp. 156–171, 2026. <https://doi.org/10.56578/ijcmem140110>.



© 2026 by the author(s). Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.

Abstract: The objective of this study was to evaluate the performance of three machine learning models for classifying and delineating critical contamination zones in a polymetallic tailings pond. Four hundred samples (water and soil) were analyzed using physicochemical variables (pH, electrical conductivity (EC), lead (Pb), and copper (Cu)). The methodology implemented Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), evaluated through 10-fold cross-validation, reporting the mean and standard deviation. The results showed that complexity is matrix-dependent: water data exhibited linear separability, allowing for perfect classification (1.0 ± 0.0), while soil data showed non-linear overlap. In this complex scenario, RF emerged as the most robust model, achieving an accuracy of 0.980 ± 0.033 and an F1-score of 0.989 ± 0.019 , surpassing the stability of SVM and KNN. It is concluded that RF is the most effective tool to minimize the risk of false negatives in spatial delimitation, guaranteeing accurate environmental remediation.

Keywords: Predictive model; Random Forest; Support Vector Machine; K-Nearest Neighbors; Machine learning; Decision frontiers

1 Introduction

Mining is currently a fundamental pillar of global economic development, serving as a major source of employment and investment in countries with abundant mineral resources. However, when it lacks responsible environmental management and a strong social commitment, it can generate significant negative impacts, most notably pollution [1].

The presence of environmental mining liabilities (EML) and generally abandoned tailings constitutes a major environmental challenge, since heavy metals can disperse into the environment due to the action of wind, surface runoff and water infiltration, which facilitates their mobilization and transport to surrounding areas [2]. Metals such as mercury (Hg), lead (Pb), arsenic (As), cadmium (Cd), and zinc (Zn) are often present in these wastes in chemically reactive forms or bound to unstable minerals, which increases their ability to dissolve and migrate in the environment [3]. This leads to its incorporation into the food chain, where it bioaccumulates and reaches toxic levels that threaten wildlife and human health [4].

In Peru, the origin of environmental liabilities (EPAs) dates back to ancestral mining activity, but their management is a contemporary challenge. The official inventory of these liabilities remains alarming. According to the latest update from the Ministry of Energy and Mines (MINEM) through Ministerial Resolution No. 338-2025-MINEM/DM, the national inventory comprises 6,122 EPAs [5].

Of that total, the La Libertad region has 135 Environmentally Disturbed Areas (PAMs) distributed across provinces such as Patá, Julcán, Otuzco, Santiago de Chuco, Sánchez Carrión, Gran Chimú, Virú, and Quiruvilca, the latter being a particularly representative case of this problem. Decades of mining activity have generated EPAs that have contributed to the contamination of the Moche River with heavy metals, causing social conflicts due to the health risks posed to nearby communities [6].

Addressing the management and evaluation of this problem requires a simultaneous analysis of multiple physicochemical variables. Pollution is not uniform but varies spatially according to a set of variables such as metal concentration, pH, and conductivity. This complexity makes it difficult to model using traditional methods, especially in scenarios characterized by nonlinear relationships and class imbalances. In this context, machine learning approaches emerge as robust tools capable of processing large volumes of multivariate data to effectively classify and delineate critical pollution zones.

Recent research has demonstrated the potential of machine learning to address classification and decision-making problems in complex systems. Bertl et al. [7] demonstrated how machine learning can transform complex rules into computational structures to generate automated and explainable decisions. In the mining field, the K-Nearest Neighbors (KNN) algorithm has been applied in various mining contexts with varying results in terms of its predictive performance. Are an example of this Kamran et al. [8], they demonstrated achieving 96% accuracy in risk classification using ISOMAP and Fuzzy C-Means. However, their effectiveness depends critically on the context. In this regard, Liu et al. [9] integrated the KNN as one of the base models in a stacking assembly scheme to evaluate the stability of excavations, where, although individually it was not the best, its collective contribution was crucial for the hybrid model to achieve an accuracy of over 92%.

For its part, the Support Vector Machine (SVM) algorithm has demonstrated high performance in various engineering applications. Wang et al. [10] reported that this model achieved exceptional performance in identifying coal and gangue, with an accuracy of 98.3%, significantly outperforming other algorithms such as KNN and Neural Networks. Huang and Zhou [11] used SVM as the base technique in a hybrid model (CS-BSVM) which, combined with assembly and optimization strategies, achieved an accuracy of 0.86 in the stability analysis of underground excavations. Liu et al. [12] implemented an optimized SVM model with the Slime Mold algorithm to predict tailings discharge volume, which demonstrated greater predictive accuracy compared to other models, with an average relative error of less than 25%, confirming its effectiveness in real engineering scenarios. Shoaib et al. [13] applied supervised models to a high-dimensional problem, concluding that SVM was the most effective algorithm (90% accuracy) for establishing optimal boundaries between different classes.

Additionally, the Random Forest (RF) algorithm has proven effective in various mining and environmental applications. Qi et al. [14] used an RF model optimized with Bayesian optimization that achieved exceptional predictive performance ($R^2 = 0.99$ in training and 0.965 in testing) to predict forms of occurrence of elements in mining waste. Şimşek et al. [15] developed an RF regression model that reliably predicted the iron grade in magnetic concentrates, achieving an R^2 of 0.735 and identifying key operating parameters through feature importance analysis. Mishra et al. [16] applied RF, along with other algorithms, to evaluate lithium tailings applications, contributing to the consensus on their technical feasibility and environmental safety, thus supporting circular economy initiatives in the mining industry.

Despite these advances, a gap exists in the literature regarding the direct comparison of these specific models for the spatial delimitation of critical zones in polymetallic tailings liabilities. While previous studies validate the individual utility of these algorithms, their comparative performance when integrating specific physicochemical parameters (pH, conductivity, Pb, and Cu) to define precise decision boundaries in complex contamination scenarios has not been sufficiently evaluated.

In this context, the present study aims to evaluate the performance of three machine learning models for the classification and spatial delimitation of critical contamination zones generated by mining EPAs in a polymetallic tailings dam located in Quiruvilca. To this end, the specific objectives are: (1) to analyze the physicochemical characteristics of the water and soil of the mining tailings, including the concentration of heavy metals (Pb and Cu), pH, and electrical conductivity (EC), for processing in ML models; (2) implement RF, SVM, and KNN models for estimating decision boundaries in the classification of critical pollution zones; (3) evaluate the performance of the models using accuracy (global precision) metrics, confusion matrices, and the precision report; and (4) compare the results obtained to identify the model with the greatest robustness in the spatial delimitation of polluted zones.

The manuscript is structured as follows: Section 2 describes the methodology, detailing the physicochemical variables (heavy metals, pH, and EC) and the implementation process of the selected predictive models (RF, SVM, and KNN). Section 3 presents the results and discussion, outlining the performance findings for each model and comparing them to identify the algorithm with the greatest accuracy in the spatial delimitation of critical zones. Finally, Section 4 presents the study's conclusions and describes future research directions.

2 Methodology

2.1 Data Collection

The research is based on the collection of a dataset obtained from water and soil samples collected directly from an abandoned polymetallic tailings dam in Quiruvilca, located in the La Libertad region, Peru. 400 samples were collected distributed at strategic points of the tailings body, both surface water and soil, which were analyzed in the laboratory following technical protocols of environmental quality. The analyses included the determination of the

parameters pH, EC, concentration of Pb, and Cu, selected for their relevance in the identification of contamination processes and acid mine drainage. The values obtained constitute the database used for the development of machine learning models, aimed at the classification and delimitation of critical areas.

2.2 Data Processing and Validation

In order to guarantee the quality and reliability of the information obtained, a process of review, filtering and validation of data was carried out before the use of the machine learning models. The records were reviewed in order to identify possible typing errors, duplicity or missing values. Which were corrected. Subsequently, the data were organized in a single table, assigning a coding to the sampling points to facilitate their processing, as well as with the parameters pH, EC, Pb, and Cu. In this way, a clean and reliable dataset was obtained, suitable for the development of predictive models (see Table 1).

Table 1. Clean data

Variables	Parameters	P1	P2	P3	P4	...	P200
Water	pH	1.88	1.79	1.90	1.84	...	6.54
	EC	10.68	12.84	10.66	11.82	...	1.05
	Pb	1.05	1.23	0.90	0.76	...	0.18
	Cu	50.06	49.19	50.62	49.17	...	0.88
Soil	pH	5.83	6.76	5.65	6.41	...	6.66
	EC	1.52	1.02	1.68	1.29	...	2.62
	Pb	100.92	95.00	98.96	99.96	...	0.12
	Cu	66.20	68.20	66.88	67.56	...	1.55

Notes: EC = electrical conductivity; Pb = lead; Cu = copper.

2.3 Preparing Data for Modeling

After completing the data cleaning, to ensure homogeneity across scales and prevent variables with larger units of measurement from dominating the process, the input features were standardized (scaled), a crucial step for the optimal performance of distance-based models such as SVM and KNN. For hyperparameter optimization, the GridSearchCV technique was applied, which works by systematically testing all possible combinations of predefined hyperparameter values on a grid [17]. Finally, the robustness and generalizability of the models were evaluated using 10-fold cross-validation, an approach that averages performance across ten subsets of data, ensuring data reliability.

2.4 Maximum Permissible Limits

The maximum permissible limit (MPL) is defined as a statistical threshold that acts as a critical boundary for classification. Test characteristics that exceed this limit are identified as a statistical outlier, labeled as potentially defective or high-risk, as it presents a significant risk of exceeding the allowable operating limits [18]. The specific MPL used in this study to determine the risk categories are detailed in Table 2.

Table 2. Maximum permissible limits (MPL)

Variables	Parameters	Units	MPL
Water	pH	-	6–9
	EC	mS/cm	1.5
	Pb	mg/L	0.4
	Cu	mg/L	1
Soil	pH	-	6–8
	EC	mS/cm	150
	Pb	mg/kg	63
	Cu	mg/kg	4

Notes: EC = electrical conductivity; Pb = lead; Cu = copper.

2.5 Machine Learning

Machine learning is a branch of AI that allows computer systems to learn from experience. Instead of relying on predefined formulas or models, machine learning algorithms employ computational methods that extract patterns directly from data. As more information is available to train the model, its performance will improve progressively

and adaptively. In supervised learning, the goal is to develop a model capable of predicting outcomes in the face of new data. This approach uses datasets in which both the expected inputs and outputs are known, so that the algorithm can be trained to generate accurate responses to unpublished information. This type of learning uses classification and regression techniques to build effective predictive models [19].

2.6 Support Vector Machine Algorithm

SVM is a supervised learning model used primarily in classification and regression problems; such application is prevalent in areas such as signal processing. The fundamental principle of this method is to find a hyperplane that allows to optimally separate different classes, in a two-dimensional space, this hyperplane is represented with a line while in a space of different dimensions it takes the shape of a plane, depending on the number of variables or characteristics analyzed. Although there may be several possible hyperplanes. SVM maximizes the separation margin, i.e. the distance between hyperplane and the nearest data points of each class [20] (Figure 1).

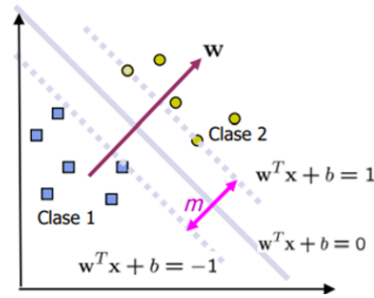


Figure 1. Decision frontier

The decision boundary is obtained from the decision function of the SVM model, which is expressed mathematically:

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

where, α_i are Lagrange's multipliers, γ_i represents the close labels, $K(x_i, x)$ is the Kernel function, and b is the term bias.

2.7 Random Forest Algorithm

RF is a versatile tool, ideal for analyzing complex data sets. This technique is applicable to both classification and regression problems, and its power lies in combining multiple decision trees. This collective approach improves the accuracy of predictions and minimizes variance error, based on statistical principles such as bagging and bootstrap [21]. In addition, it stands out as an evolution of classical statistics that enhances the capabilities of technological mechanisms. To fully understand it, it is essential to familiarize yourself with the concept of bootstrap sampling, the foundation on which this model is built [22]. In regression problems, RF is calculated through the average of the predictions of each decision tree in the forest [23].

$$\hat{y}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (2)$$

where, B is the total number of trees, $T_b(x)$ is the value predicted by the b -tree for input x .

2.8 K-Nearest Neighbors Algorithm

KNN, despite its conceptual simplicity, is recognized as a fundamental technique of local regression as shown in Figure 2. Their logic is based on the assumption that, in a small, local region of the data space (the nearest neighbors), the conditional mean, which represents the optimal prediction, remains approximately constant [24].

It is a supervised learning model that operates under the principle of proximity: it assumes that data is nearby. To predict a new point, the model identifies its nearest K neighbors of the training set, as measured by a distance metric (Euclidean). The final prediction is determined by majority voting (the most frequent class) for classification or by the average of the neighbors' values for regression. The Euclidean distance is calculated with the following formula:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

where, x_i are the coordinates of points in x-axis space, y_i are the coordinates of points in y-axis space and n is the number of dimensions.

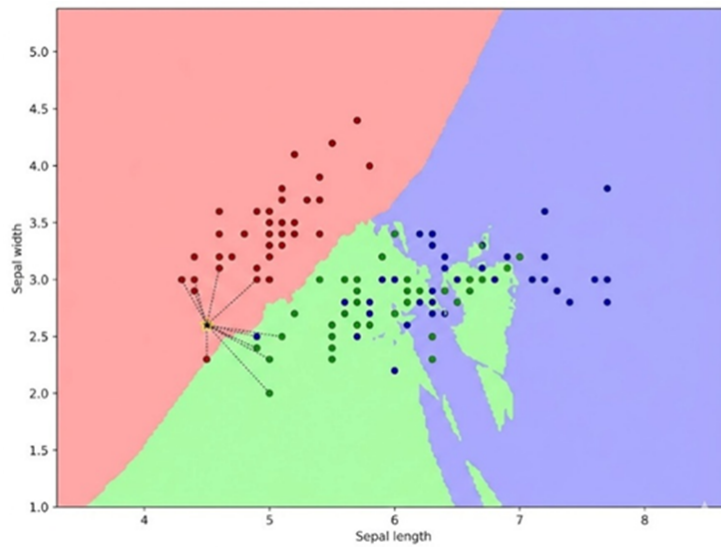


Figure 2. K-Nearest Neighbors (KNN) graphic representation

2.9 Performance Metrics

2.9.1 Accuracy

It is one of the most widely used measures to evaluate the performance of classification models in machine learning. Its main function is to determine the overall accuracy of the model, i.e. how many predictions made match the actual values. This metric is expressed by the formula:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (4)$$

In addition, it provides a general measure of the accuracy of a classification model, indicating how often predictions match real classes. However, its interpretation may be limited when there is an imbalance in the distribution of classes, since a model could obtain high accuracy by predicting predominantly the majority class, without reflecting balanced performance in all categories [25].

2.9.2 Confusion matrix

The confusion matrix is a critical tool for evaluating the performance of classification models in machine learning. It is represented as a square table of dimension $N \times N$, where N corresponds to the number of classes or categories in the model. In it, the columns indicate the actual values, and the rows show the values predicted by the model. Each cell reflects the number of instances ranked in a specific combination of real and predicted classes, allowing Distinguish between successes and errors. In its simplest form (binary classification), the matrix is composed of four elements: true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) [26].

2.9.3 Accuracy report

The accuracy report is a statistical tool used to evaluate the performance of classification models, showing key metrics derived from the confounding matrix, such as accuracy, recall, and F-measure. According to Powers [27] accuracy is defined as the proportion of cases predicted as positive that actually are, expressed by the formula:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

where, TP are the true positives and FP are the false positives.

This metric reflects the degree of confidence that can be had in the positive predictions of the model. For its part, recall measures the model's ability to correctly identify all real positive cases, while F1 combines both metrics through its harmonic average, providing a balance between accuracy and sensitivity. Powers warns that these measures, while useful, can be biased if they are applied without considering the level of prevalence of classes or chance, as they ignore correctly classified negative cases.

2.10 Procedure

The procedure is presented in Figure 3. This figure details the steps that will be followed during the research process with the aim of evaluating machine learning models for the classification and spatial delimitation of critical areas of contamination in a polymetallic tailings dam.

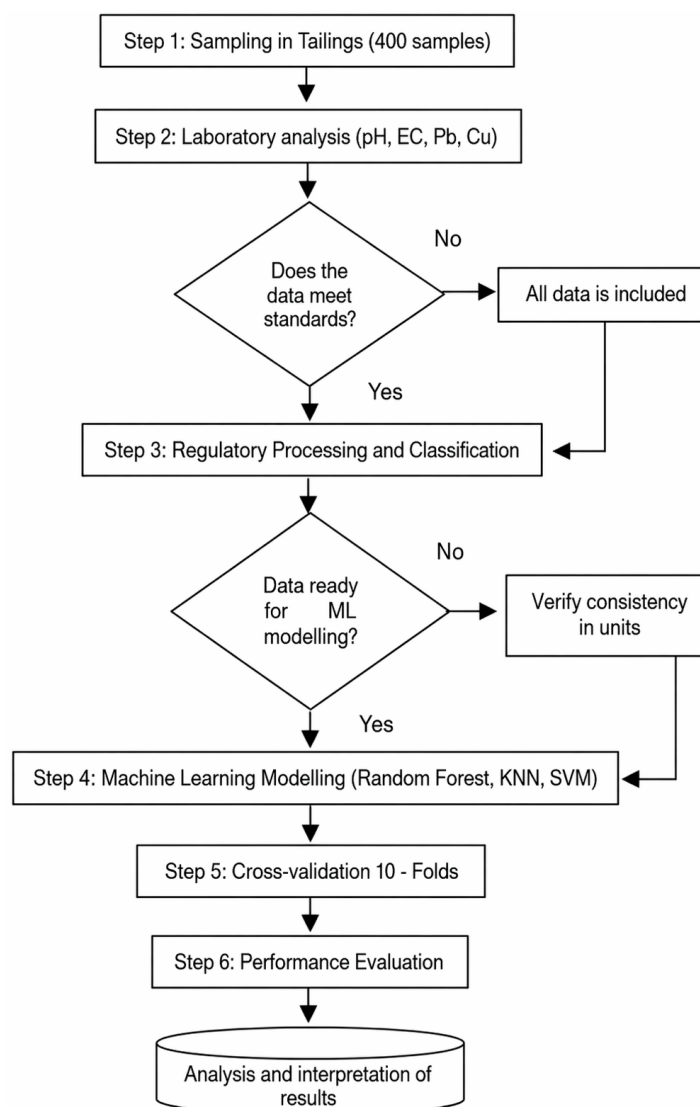


Figure 3. Research flowchart

2.11 Population

The study population is made up of the total number of sampling points (water and soil) with their corresponding physicochemical characteristics (concentration of heavy metals, pH, EC) within the area of direct influence of the abandoned polymetallic tailings dam in Quiruvilca.

2.12 Sample

In this study, the total dataset consists of 400 records collected directly from the tailings pond (200 water and 200 soil). Since the objective is predictive modeling using machine learning, all available samples were used as the study population to ensure a robust representation of the site's physicochemical variability. This sample size is sufficient for the convergence of the proposed classification algorithms and cross-validation of 10 folds.

2.13 Study Design

This study is framed in a quantitative approach, since it allows the analysis of numerical data for the classification of pollution areas. Variables of interest, such as metal concentration (Pb and Cu), pH, and EC, are statistically

measured and processed. A comparative and evaluative design was chosen in order to implement, train and measure the performance of three machine learning algorithms (RF, SVM and KNN) to determine which one offers the highest precision in the delimitation of critical pollution zones. To guarantee the validity of the results and the robustness of the predictive models, the 10-Fold Cross-Validation technique was implemented, which consists of dividing the data into 10 stratified parts, using nine parts to train and one to validate, until calculating the average accuracy that will determine the most reliable model [28].

3 Results

Following the methodology and objectives of the research, this section presents the performance of the RF, SVM, and KNN models.

3.1 Data Processing

To ensure the accuracy and robustness of the prediction models, the 400 collected records (200 water and 200 soil) were processed and characterized. A normative classification was performed, labeling each sample as Class 0 (Uncontaminated) or Class 1 (Contaminated), applying the MPLs for our variables. For this integration, a strict environmental safety criterion was adopted: a sample was labeled as Class 1 if at least one of the four parameters (pH, EC, Pb, or Cu) exceeded its respective MPL. Conversely, the Class 0 label was assigned only when all analyzed parameters were within the permitted ranges. Furthermore, K-fold cross-validation (10 folds) was used during training to ensure that the models could accurately generalize to unseen data, guaranteeing the reliability of the findings.

3.2 Defining Input and Output Parameters for Predictive Models

Four input variables were defined, which consist of the physicochemical characteristics analyzed in the 400 samples. The output variable is the “Class”, which is a categorical value determined by the normative classification of the MPL (see Table 3).

Table 3. Variable selection

Variables	Cant.	Guy	Input/Output
pH	400	float	Input
Electrical conductivity (EC)	400	float	Input
Lead (Pb)	400	float	Input
Copper (Cu)	400	float	Input
Class	400	integer	Output

3.3 Exploratory Analysis and Normative Classification

By applying the MPL (Table 4) to each dataset, I yield two distinct modeling scenarios, each representing a major challenge for the algorithms. The classification of the water samples resulted in a dataset with a balance of 150 contaminated samples (Class 1) and 50 uncontaminated samples (Class 0). This set represents a high separability, constituting a basic challenge, to test the ability of the models to achieve optimal accuracy in a controlled and low complexity scenario. On the other hand, the classification of the soil samples revealed a highly complex scenario. The result was a severe class imbalance, yielding 174 contaminated samples (Class 1) and 26 uncontaminated samples (Class 0), this represents an advanced challenge, since the models must handle a high nonlinear overlap in the data and must correctly identify the minority class, testing the robustness and optimization of each algorithm.

Table 4. Class balance

Tuition	Water Samples	Soil Samples
0	50	26
1	150	174

This difference in the nature of the water data is confirmed in the exploratory analysis as seen in Figure 4. Scatter plots (Pair Plots) show two clearly defined groups of points (Class 0 and Class 1) occupying distinct regions in the feature space.

As shown in Figure 4, a complete scatter plot is presented, visualizing the pairwise relationships between all the physicochemical variables. Taking the intersection of pH (x-axis) and EC (y-axis) as a specific example, the

separation is complete: the red dots (Class 1, Contaminated) form a clear cluster in the upper left corner, indicating low pH (acidic) and high conductivity. Conversely, the blue dots (Class 0, Uncontaminated) are clustered in the lower right corner (basic pH, low conductivity). This arrangement in opposite corners visually confirms that, for water, there is a scenario of low complexity and high linear separability.

The box plots in Figure 5 strongly reinforces this conclusion. They demonstrate that there is complete separation between the classes for all four variables analyzed. The interquartile ranges (the boxes) of Class 0 and Class 1 do not overlap. A clear dichotomy is observed, where the uncontaminated samples (Class 0, blue) consistently show a high pH and low levels of EC, Pb and Cu, while the contaminated samples show an opposite pattern; very low pH and drastically higher levels of EC, Pb and Cu.

In contrast, the scatter plots (pair plots) of the soil samples show a clear complexity as can be seen in Figure 6. There is an imbalance of Classes, because in the diagonal, the red density curves (Class 1) are massive, while the blue curves (Class 0) are tiny, in addition there is a high overlapping (Overlapping), in all scatters plots the blue dots are mixed within the swarm of red dots.

The box plots (Figure 7) reinforce this conclusion of the high complexity of the soil. Unlike the water graphs, where the boxes were properly separated, here the ranges of the Class 0 (blue) and Class 1 (red) boxes overlap significantly in almost all variables, especially in EC, Pb and Cu. This overlay indicates a lack of a clear separator, for example, the EC graph shows that the Class 0 range is almost completely contained within the Class 1 range. This means that no single variable is sufficient to reliably classify the data.

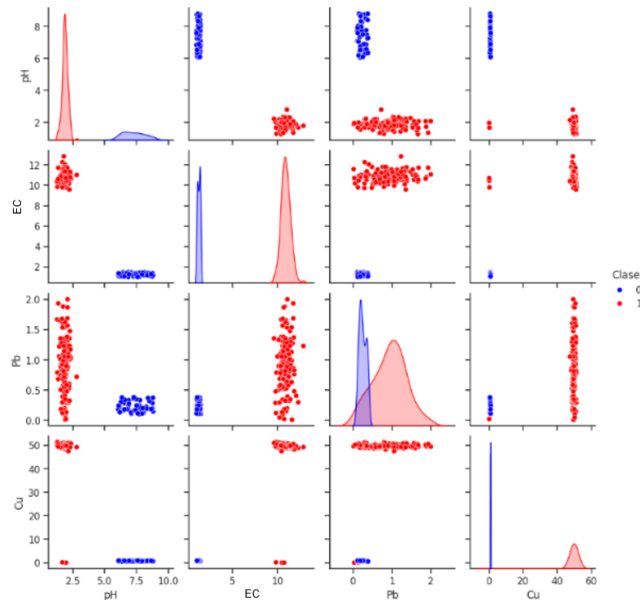


Figure 4. Scatter plot by class (water)

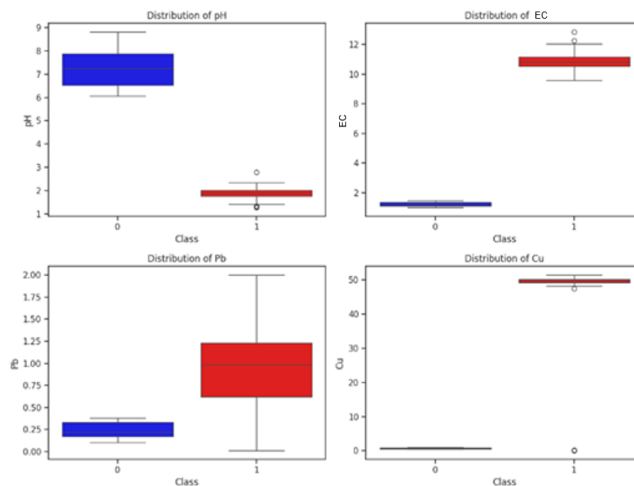


Figure 5. Distribution of variables by classes (water)

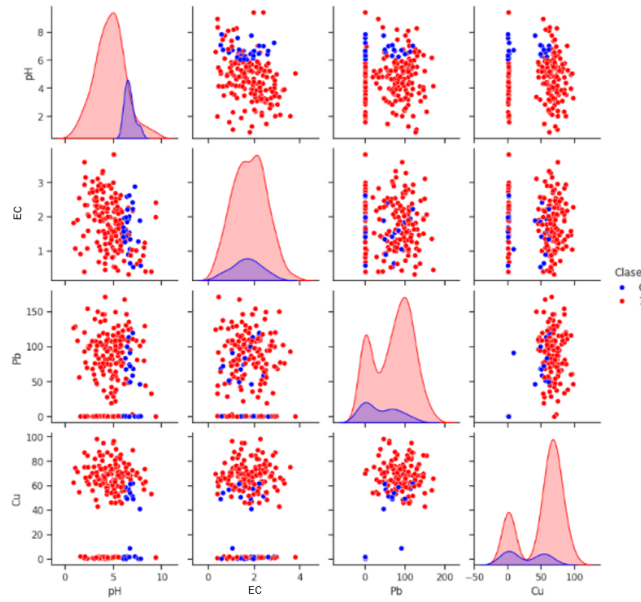


Figure 6. Scatter plot by class (soil)

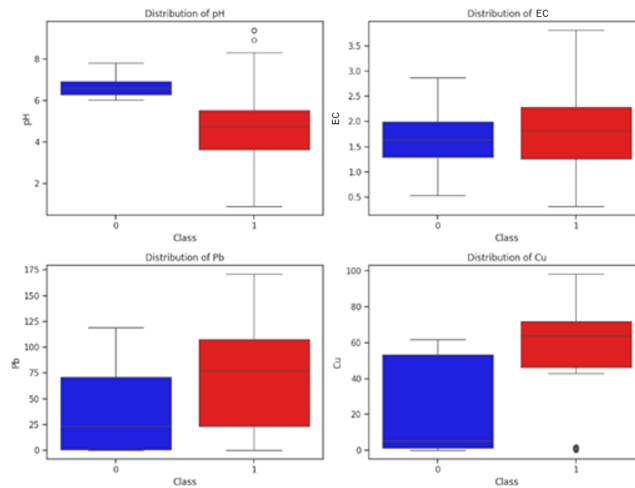


Figure 7. Distribution of variables by classes (soil)

3.4 Training and Validation of Predictive Models

To address the challenges of each dataset, a hyperparameter optimization (tuning) process was implemented for each of the trained models (RF, SVM, and KNN) using GridSearchCV. Furthermore, to ensure robust analysis in the soil sample imbalance scenario, the `class_weight = balanced` parameter was enabled in RF, and the “F1-Score (Macro)” metric was used for optimization in SVM. The performance of the optimized models was validated using 10-fold cross-validation.

All three models in the water sample challenge achieved perfect performance, as shown in Table 5, reaching 100% in all evaluation metrics. This suggests high data separability, as evidenced in the visual exploration (Figure 4 and Figure 5). To mathematically confirm this observation and rule out overfitting, further experimental validation was performed by implementing a linear kernel SVM. This simple model also achieved absolute accuracy (1.0 ± 0.0), demonstrating that the complexity of nonlinear models is redundant for this specific matrix. Detailed results for all models are presented in Table 5.

On the other hand, significant differences emerged in the soil data scenario. Although optimization improved overall performance, RF demonstrated clear superiority in both accuracy and stability. Table 6 details that RF obtained the highest Accuracy (0.980 ± 0.033) and the highest F1-Score (0.989 ± 0.019). This performance surpasses that of SVM (Accuracy: 0.965 ± 0.050 ; F1-Score: 0.980 ± 0.030) and KNN (Accuracy: 0.960 ± 0.037 ; F1-Score: 0.976

± 0.021). Considering that the F1-Score is the most critical metric for class imbalance, and observing the lower dispersion in the RF results, it is confirmed as the most robust and balanced model.

Table 5. Performance metrics (water)

Model	Random Forest (RF)	Support Vector Machine (SVM)	K-Nearest Neighbors (KNN)	SVM (Linear)
Accuracy	1.0 \pm 0.0	1.0 \pm 0.0	1.0 \pm 0.0	1.0 \pm 0.0
Precision	1.0 \pm 0.0	1.0 \pm 0.0	1.0 \pm 0.0	1.0 \pm 0.0
Recall	1.0 \pm 0.0	1.0 \pm 0.0	1.0 \pm 0.0	1.0 \pm 0.0
F1-Score	1.0 \pm 0.0	1.0 \pm 0.0	1.0 \pm 0.0	1.0 \pm 0.0

Note: 1. Values represent the mean \pm standard deviation obtained through 10-fold cross-validation. 2. SVM (Radial Basis Function): standard model optimized using the Radial Basis Function kernel for nonlinear pattern detection. 3. SVM (Linear): additional implementation model with a linear kernel to empirically verify the linear separability hypothesis in water samples.

Table 6. Performance metrics (soil)

Model	Random Forest (RF)	Support Vector Machine (SVM)	K-Nearest Neighbors (KNN)
Accuracy	0.980 \pm 0.033	0.965 \pm 0.050	0.960 \pm 0.037
Precision	0.989 \pm 0.033	0.988 \pm 0.035	0.978 \pm 0.037
Recall	0.989 \pm 0.021	0.973 \pm 0.036	0.976 \pm 0.030
F1-Score	0.989 \pm 0.019	0.980 \pm 0.030	0.976 \pm 0.021

3.5 Analysis of Results

The results in section 3.4 already presented a clear contrast in model performance. While all three optimized models achieved 100% perfect performance with water samples (Table 5), the metrics for soil sample results (Table 6) identified RF as the superior model (98.0%), ahead of SVM (96.5%) and KNN (96.0%). Analyzing the confusion matrices and decision boundaries is crucial to understanding why the SVM and KNN models underperformed with soil samples and the practical implications of their errors.

As expected from the perfect metric results (100%), the confusion matrices (Figure 8) confirm that all three models (RF, SVM, and KNN) classified all 200 water samples without error.

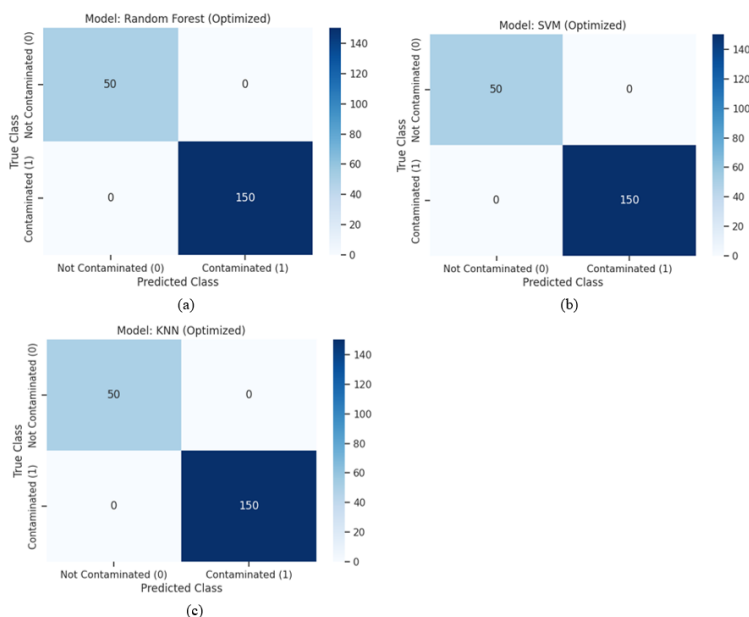


Figure 8. Confusion matrices (water): (a) Random Forest; (b) Support Vector Machine; (c) K-Nearest Neighbors

Through the decision boundaries (Figure 9) it is visualized how the three models draw clear boundaries that successfully separate the “Uncontaminated” samples from the “Contaminated” ones, validating their performance capacity in a low-complexity scenario.

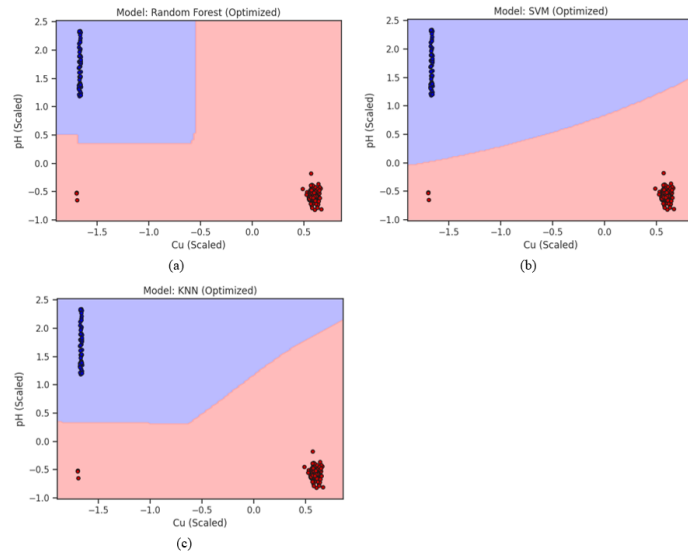


Figure 9. Decision boundaries (water): (a) Random Forest; (b) Support Vector Machine; (c) K-Nearest Neighbors

The analysis of the soil scenario reveals the true complexity of the problem. Unlike with water, the cross-validation metrics (Table 6) indicate that no model achieved absolute perfection, which is consistent with the nonlinear overlap of the classes. However, the confusion matrices confirm the superiority of RF, which significantly minimized classification errors, achieving the greatest stability (F1-Score: 0.989 ± 0.019). In contrast, the SVM and KNN models exhibited higher error rates (Accuracy: 0.965 and 0.960, respectively). Specifically, SVM struggled to identify all critical samples, producing four false negatives by misclassifying four contaminated samples (Class 1) as uncontaminated (Class 0), resulting in lower sensitivity (Recall: 0.973) compared to the adaptability of RF (Recall: 0.989).

The confusion matrix for KNN (Figure 10c) shows an absence of classification errors in the critical class because its decision criterion based on Euclidean proximity allows it to adapt to local islands of contaminated data without being forced to generalize a rigid global boundary. However, although KNN matched RF in identifying positives in this instance, its larger standard deviation in cross-validation (Table 6) suggests that its accuracy depends more on the specific distribution of the training data than that of RF.

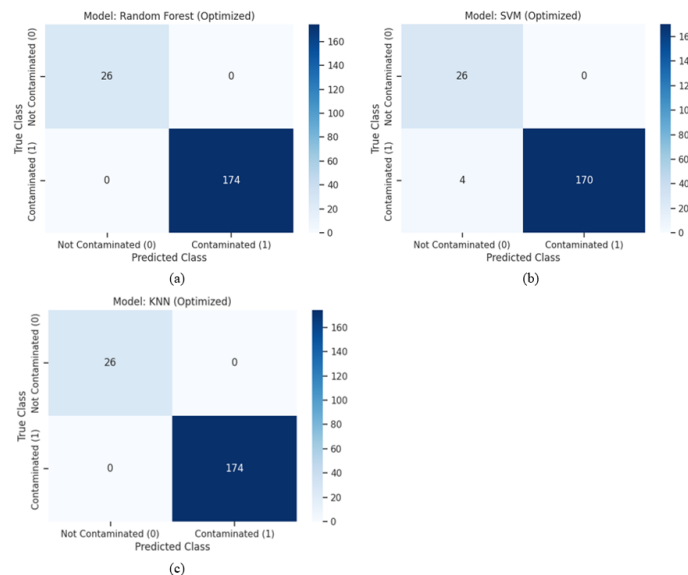


Figure 10. Confusion matrices (soil): (a) Random Forest; (b) Support Vector Machine; (c) K-Nearest Neighbors

The reason for this SVM failure is analytically explained by the nature of its optimization: by attempting to maximize the separation margin with a smooth boundary (Figure 11b), the SVM is unable to circumvent irregular intrusions in the minority class, losing sensitivity in overlapping areas. In contrast, locality- and partition-based

algorithms, such as KNN and RF, demonstrated a superior ability to model these irregularities.

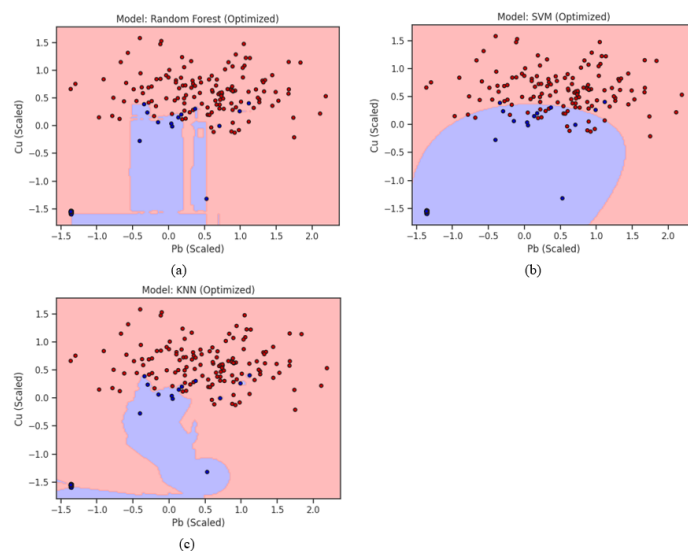


Figure 11. Decision boundaries (soil): (a) Random Forest; (b) Support Vector Machine; (c) K-Nearest Neighbors

It should be noted that visualizing decision boundaries involves dimensional simplification, as it is impossible to graph the real hyperplane in a 4-dimensional space. Therefore, the pairs of variables with the most important characteristics were selected for each matrix. In the case of water (Figure 9), pH and Cu were projected, given that extreme acidity (pH) is the strongest physical discriminant in acid drainage, allowing for a clear linear separation. On the other hand, for soil samples (Figure 11), where pH is less decisive, Pb and Cu were selected, since the concentration of these heavy metals constitutes the dominant predictive factor in the solid matrix. This distinction clearly illustrates the models’ strategy: while separation is trivial in water, in soil the inability of the smooth SVM boundary to handle the complex overlap between metals is evident, in contrast to the precise, nonlinear boundaries that RF and KNN manage to define.

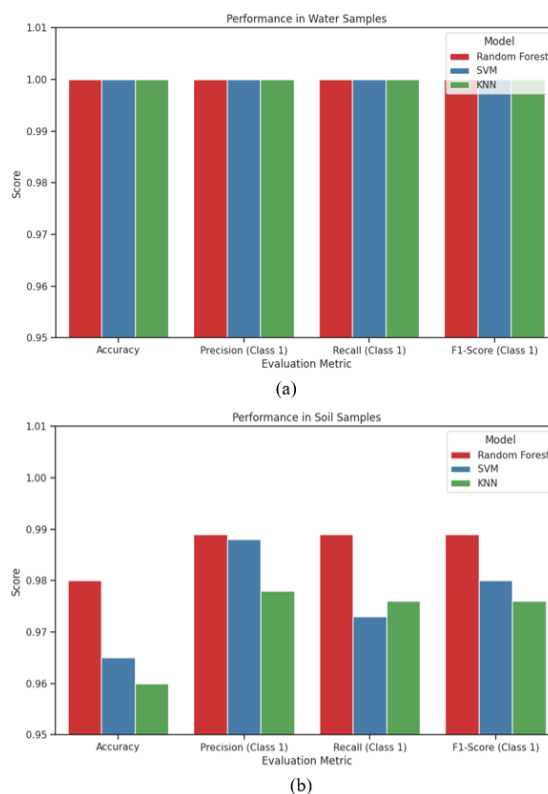


Figure 12. Comparative performance bar charts: (a) Model performance (water); (b) Model performance (soil)

This section concludes with a visual summary of the performance obtained in both scenarios, allowing for immediate identification of the optimal model. Figure 12 presents a graphical comparison of the key metrics (Accuracy, Precision, Recall, and F1-Score) for the optimized models on the water and soil datasets.

The performance graph of the models with water samples (Figure 12a) demonstrates the ability of all three algorithms to solve a basic challenge, showing perfect 100% performance across all metrics. In contrast, the difference in height of the soil bars (Figure 12b) magnifies the performance disparity. RF's F1-Score is the highest (0.989 ± 0.019), confirming its role as the most robust model. While KNN showed competitive performance, its overall Accuracy (0.960 ± 0.037) and F1-Score (0.976 ± 0.021) metrics place it below RF in terms of predictive power and stability. This numerical difference validates RF as the essential tool for ensuring spatial delimitation of critical areas with maximum precision and the lowest risk of error.

4 Discussion

4.1 Discussion of Results

The superiority of RF is not a coincidence, but a direct validation of its inherent capabilities. The model's excellent performance (F1-Score: 0.989 ± 0.019) aligns with previous research, such as that by Qi et al. [14], who demonstrated RF's ability to achieve exceptional predictions in complex environments. Our study confirms this, showing how the RF architecture is capable of creating nonlinear boundaries (Figure 11a) that effectively resolve the high class overlap in the ground.

Furthermore, although authors such as Shoaib et al. [13] highlighted the effectiveness of SVM in high-dimensional problems, our analysis demonstrates that, in the specific context of severe imbalance and fuzzy boundaries, this model presents critical limitations. A forensic analysis of the four samples that SVM failed to detect (False Negatives) reveals that they share a distinctive characteristic: they are "borderline" instances (borderline cases) with high physicochemical ambiguity. These samples exhibit Pb and Cu concentrations that only marginally exceed the MPLs, while their pH and EC values are similar to those of the uncontaminated group. Because SVM optimization prioritizes maximizing the overall margin to ensure a smooth boundary, the algorithm tended to treat these local points as noise, incorrectly classifying them as safe. In contrast, RF's ability to segment the space using hierarchical decision rules allowed it to isolate and correctly classify these subtle anomalies, validating the need for robust tools for environmental safety, as suggested by Bertl et al. [7].

4.2 Practical Application and Implementation Considerations

The superiority demonstrated by RF translates into a clear benefit of cost optimization and accuracy in the field. The model's robustness allows for precise geospatial delineation, which is essential for distinguishing contaminated from uncontaminated areas. A crucial benefit is the reduction of costs associated with false negatives, eliminating the risk of leaving critical contamination zones untreated. This prevents the escalation of risks and the high future costs associated with delayed remediation.

The successful implementation of this predictive model requires overcoming operational challenges and establishing a suitable technological infrastructure. Furthermore, robust data collection systems using sensors or standardized protocols are essential for quickly obtaining input parameters. Finally, training personnel is also fundamental so they understand and trust the predictive tool, ensuring effective management and its scalability to other areas of mining liabilities.

5 Conclusions

5.1 Key Findings, Implications and Contributions

The training and evaluation of three machine learning models (RF, SVM, and KNN) for classifying contamination zones in a polymetallic tailings pond yielded the following results.

The complexity of the classification problem depends on the analyzed matrix (Water vs. Soil). The analysis of water data revealed a simple and linearly separable problem, while the soil data presented a complex, nonlinear problem with a high degree of class overlap.

The RF model emerged as the most robust and reliable algorithm for complex scenarios, maintaining superior predictive performance with an Accuracy of 0.980 ± 0.033 and an F1-Score of 0.989 ± 0.019 in the soil data. This performance significantly surpasses that of SVM (0.965 ± 0.050 Accuracy) and KNN (0.960 ± 0.037 Accuracy). In contrast to the water data, where all models achieved excellent accuracy of 100%.

In terms of applicability, the performance difference between the models is crucial. While the use of suboptimal algorithms such as KNN or SVM would imply assuming a latent error rate and lower sensitivity to contaminants, RF proved to be the only tool capable of guaranteeing the necessary stability (Recall: 0.989 ± 0.021) for safe remediation. Consequently, the main contribution of this study is not only the metric comparison, but also the validation that ensemble learning models are structurally superior to distance- or margin-based methods for resolving the nonlinear complexity of mining-related EPAs.

5.2 Limitations of the Study and Future Lines of Research

This study has certain limitations. First, the analysis is restricted to a set of 400 samples (200 water and 200 soil) from a single polymetallic tailings dam (Quiruvilca), which may limit the generalization of the results to other mines with different geological conditions. Second, the predictive model was based on four characteristics (pH, EC, Pb, and Cu); The inclusion of other relevant heavy metals could alter the behavior of the model.

Future lines of research focus on validating the robustness of this methodology by applying the RF model in other tailings dams in Peru. It is also suggested that the dataset be expanded to include a broader spectrum to include variables, physicochemical. Finally, this RF model could be adapted for integration into automated monitoring platforms, allowing real-time classification of contamination and optimizing the efficiency of closure and environmental remediation plans.

Author Contributions

Conceptualization, D.H. and M.H.; methodology, D.H., M.H., and J.N.; software, J.N.; validation, E.N. and W.E.; formal analysis, D.H., M.H., and J.N.; investigation, D.H., M.H., and J.N.; resources, W.E.; data curation, E.N.; writing—original draft preparation, D.H., M.H., and J.N.; writing—review and editing, E.N. and J.N.; visualization, W.E.; supervision, E.N. and W.E.; project administration, E.N.; funding acquisition, W.E. All authors have read and agreed to the published version of the manuscript.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] I. A. L. Eyzaguirre, “(In) sustainability of mining in Latin America? A systematic review of trends and gaps,” *Rev. Dep. Geogr.*, vol. 42, p. e190162, 2022. <https://doi.org/10.11606/eISSN.2236-2878.rdg.2022.190162>
- [2] W. Li, Y. Deng, H. Wang, Y. Hu, and H. Cheng, “Potential risk, leaching behavior and mechanism of heavy metals from mine tailings under acid rain,” *Chemosphere*, vol. 350, p. 140995, 2024. <https://doi.org/10.1016/j.chemosphere.2023.140995>
- [3] D. Castillo, K. Palma, L. Santander, H. Bolaños, G. Palma, and P. Navarro, “Physical, chemical, mineralogical, and toxicological characterization of active and inactive tailings in the Arequipa Region, Peru,” *Minerals*, vol. 15, no. 8, p. 830, 2025. <https://doi.org/10.3390/min15080830>
- [4] J. P. R. de Souza, J. Garnier, J. M. Quintarelli, M. de Sousa Tonhá, H. L. Roig, P. Seyler, and J. R. de Souza, “Adapted sequential extraction protocol to study mercury speciation in gold mining tailings: Implications for environmental contamination in the Amazon,” *Toxics*, vol. 12, no. 5, p. 326, 2024. <https://doi.org/10.3390/toxics12050326>
- [5] Ministry of Energy and Mines, “Ministerial resolution No. 338-2025-MINEM/DM: Approval of the update of the initial inventory of mining environmental liabilities,” 2025. <https://cdn.www.gob.pe/uploads/document/file/7667900/6496380-rm-n-056-2025-minem-dm.pdf?v=1740097571>
- [6] A. Delgado, F. Martínez, J. Villanueva, A. Torres, and L. Andrade-Arenas, “Model based on grey systems to assess water quality from mines in operation and environmental liabilities,” *Int. J. Eng. Trends Technol.*, vol. 70, no. 9, pp. 346–353, 2022. <https://doi.org/10.14445/22315381/IJETT-V70I9P234>
- [7] M. Bertl, S. Price, and D. Draheim, “Transforming legal texts into computational logic: Enhancing next generation public sector automation through explainable AI decision support,” *Int. J. Cogn. Comput. Eng.*, vol. 7, pp. 40–57, 2026. <https://doi.org/10.1016/j.ijcce.2025.07.003>
- [8] M. Kamran, B. Ullah, M. Ahmad, and M. M. S. Sabri, “Application of KNN-based isometric mapping and fuzzy c-means algorithm to predict short-term rockburst risk in deep underground projects,” *Front. Public Health*, vol. 10, p. 1023890, 2022. <https://doi.org/10.3389/fpubh.2022.1023890>
- [9] L. Liu, G. Zhao, W. Liang, and Z. Jian, “Hybrid stacking ensemble algorithm and simulated annealing optimization for stability evaluation of underground entry-type excavations,” *Undergr. Space*, vol. 17, pp. 25–44, 2024. <https://doi.org/10.1016/j.undsp.2023.11.002>
- [10] X. Wang, M. Wu, and P. Zhao, “Dielectric identification method and system design of coal gangue based on frequency shift characteristics,” *Sci. Rep.*, vol. 15, no. 1, p. 8712, 2025. <https://doi.org/10.1038/s41598-025-93401-x>

- [11] S. Huang and J. Zhou, “An enhanced stability evaluation system for entry-type excavations: Utilizing a hybrid bagging-SVM model, GP and kriging techniques,” *J. Rock Mech. Geotech. Eng.*, vol. 17, no. 4, pp. 2360–2373, 2025. <https://doi.org/10.1016/j.jrmge.2024.05.024>
- [12] G. Liu, B. Zhao, X. Kong, Y. Xin, M. Wang, and Y. Zhang, “A SMA-SVM-based prediction model for the tailings discharge volume after tailings dam failure,” *Water*, vol. 17, no. 4, p. 604, 2025. <https://doi.org/10.3390/w17040604>
- [13] M. Shoaib, M. U. Afridi, and G. Husnain, “Forecasting algorithm for the vocational mobility of graduates with multiple attributes,” *Expert Syst. Appl.*, vol. 297, p. 129463, 2026. <https://doi.org/10.1016/j.eswa.2025.129463>
- [14] C. Qi, T. Hu, J. Zheng, K. Li, N. Zhou, M. Zhou, and Q. Chen, “Artificial intelligence-based prediction model for the elemental occurrence form of tailings and mine wastes,” *Environ. Res.*, vol. 249, p. 118378, 2024. <https://doi.org/10.1016/j.envres.2024.118378>
- [15] S. Şimşek, Y. Cebeci, and A. Abbaker, “Sustainable iron ore tailings reprocessing: Machine learning prediction and integrated beneficiation strategies,” *Can. Metall. Q.*, pp. 1–15, 2025. <https://doi.org/10.1080/00084433.2025.2550787>
- [16] A. Mishra, P. Vedadi, N. Binte, A. Fosu, Q. Dehaine, P. Heikkilä, N. Kanari, H. Chambart, M. Räisänen, and P. Perumal, “Exploring the potential of lithium tailings in construction materials,” *Miner. Eng.*, vol. 233, p. 109678, 2025. <https://doi.org/10.1016/j.mineng.2025.109678>
- [17] M. S. Sawah, H. Elmannai, A. A. El-Bary, K. Lotfy, and O. E. Sheta, “Distributed denial of service (DDoS) classification based on random forest model with backward elimination algorithm and grid search algorithm,” *Sci. Rep.*, vol. 15, no. 1, p. 19063, 2025. <https://doi.org/10.1038/s41598-025-03868-x>
- [18] Y. Koo, W. Shin, H. Park, H. Yang, and C. Nam, “Analysis of latent defect detection using sigma deviation count labeling (SDCL),” *Electronics*, vol. 14, no. 19, p. 3912, 2025. <https://doi.org/10.3390/electronics14193912>
- [19] A. Géron, *Aprende Machine Learning con Scikit-Learn, Keras y TensorFlow: Conceptos, Herramientas y Técnicas Para Conseguir Sistemas Inteligentes*. ANAYA Multimedia, 2023.
- [20] G. A. Betancourt, “Las máquinas de soporte vectorial (SVMs),” *Sci. Tech.*, vol. 10, no. 27, pp. 67–72, 2005.
- [21] A. F. Alaminos-Fernández, *Árboles de decisión en R con Random Forest*. Alicante: Limencop, 2022.
- [22] O. Theobald, *Machine Learning for Absolute Beginners*. Scatterplot Press, 2021.
- [23] A. A. Khan, O. Chaudhari, and R. Chandra, “A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation,” *Expert Syst. Appl.*, vol. 244, p. 122778, 2024. <https://doi.org/10.1016/j.eswa.2023.122778>
- [24] R. Fernández Casal, J. Costa, and M. Oviedo de la Fuente, *Métodos Predictivos de Aprendizaje Estadístico*. Universidade da Coruña, Servizo de Publicacións, 2024. <https://doi.org/10.17979/spudc.9788497498937>
- [25] A. Zapeta Hernández, G. A. Galindo Rosales, H. J. Juan Santiago, and M. Martínez Lee, “Performance metrics to evaluate machine learning in oil image classification using convolutional neural networks,” *Cienc. Latina Rev. Cient. Multidiscip.*, vol. 6, no. 5, pp. 4624–4637, 2022. https://doi.org/10.37811/cl_rcm.v6i5.3420
- [26] Z. Karimi, “Confusion matrix,” 2021. <https://www.researchgate.net/publication/355096788>
- [27] D. M. W. Powers, “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation,” *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 2229–3981, 2011. <https://doi.org/10.9735/2229-3981>
- [28] Y. Rimal, N. Sharma, S. Paudel, A. Alsadoon, M. P. Koirala, and S. Gill, “Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy,” *Sci. Rep.*, vol. 15, no. 1, p. 13444, 2025. <https://doi.org/10.1038/s41598-025-93675-1>

Nomenclature

pH	Hydrogen potential
EC	Electrical conductivity, $mS \cdot cm^{-1}$ (water/soil)
Cu	Copper concentration, $mg \cdot L^{-1}$ (water)/ $mg \cdot kg^{-1}$ (soil)
Pb	Lead concentration, $mg \cdot L^{-1}$ (water)/ $mg \cdot kg^{-1}$ (soil)
MPL	Maximum permissible limit
\hat{y}	Predicted value
N_t	Number of trees in random forest
x, y	Coordinates or features of data samples
α_i	Lagrange’s multipliers

γ_i	Close labels
$K(x_i, x)$	Kernel function
b	Term bias
x_i	Coordinates of points in X-axis space
TP	True positives
FP	False positives
TN	True negatives
FN	False negatives
n	Sample size
N	Population size
Z	Confidence level
E	Error rate
S	Standard deviation

Greek symbols

Σ	Summation operator
μ	Mean value (in feature standardization)
σ	Standard deviation (in data scaling)
Δ	Difference or variation (used in accuracy/error expressions)

Subscripts

0	Class 0—uncontaminated sample
1	Class 1—contaminated sample
train/test	Training and testing subsets in cross-validation
RF/SVM/KNN	Refers to model in metrics or figures