Shahrood University of Technology

Journal of Mining and Environment (JME)

Journal homepage: www.jme.shahroodut.ac.ir

Iranian Society of Mining Engineering (IRSME)

Multimodel comparison of supervised algorithms for lithological classification in a gold deposit in Peru

Marco Antonio Cotrina-Teatino* and Jairo Jhonatan Marquina-Araujo

Department of Mining Engineering, Faculty of Engineering, National University of Trujillo, Trujillo, Peru

Article Info

Received 1 December 2025

Received in Revised form 27 December 2025

Accepted 6 February 2026

Published online 6 February 2026

DOI: 10.22044/jme.2026.17252.3415

Keywords

Lithological classification

Machine learning

Gold mining

Uncertainty

Abstract

Integrating entropy-based uncertainty analysis with machine learning offers a novel approach to improving lithological classification in mineral exploration. This study applies supervised algorithms to predict lithology from spatial and geochemical data collected at a gold deposit in northern Peru. The dataset includes 2,129 composited samples from 140 drillholes, containing spatial coordinates (East, North, Elevation) and gold content (Au). Six classifiers were tested: Random Forest, XGBoost, Support Vector Machine, K-Nearest Neighbors, Naive Bayes, and Multilayer Perceptron. Stratified five-fold cross-validation was applied to a 70/30 train-test split. The best performance was achieved by ANN-MLP (94.5% accuracy) and XGBoost (93.9%), with F1-scores above 94%. In zones of low uncertainty, models reached up to 100% precision, while accuracy dropped to 71.9% in highly uncertain regions. Entropy-based uncertainty mapping highlighted areas of geological ambiguity, such as lithological boundaries or sparsely sampled zones. The Friedman test confirmed statistically significant differences among classifiers ($p < 0.001$). These findings demonstrate that combining machine learning with spatial uncertainty quantification enhances both predictive reliability and geological interpretability, offering a practical tool for guiding exploration and reducing risk in complex mineral systems.

1. Introduction

Accurate lithological classification is critical in mineral resource modeling and mine planning, as it directly affects orebody delineation, blast optimization, and grade control. However, in structurally complex or stratigraphically heterogeneous deposits, such as orogenic gold systems determining lithological contacts and transitions remains a challenge, due to the ambiguous nature of rock boundaries and the limited resolution of drilling data [1, 2]. This complexity often leads to misclassification and reduced confidence in block models, ultimately affecting operational and economic decisions.

Traditionally, lithology is identified through core logging, cuttings analysis, and geophysical surveys. While core analysis is among the most accurate techniques, its application is limited by cost and time, often leading to discontinuous datasets that undermine the resolution and continuity of geological models [3–5].

Machine learning (ML) algorithms have proven effective for lithological classification using data from drilling operations. Support Vector Machines (SVM) have been successfully employed to predict rock mass properties and classify heterogeneous lithologies with promising results [6–8]. Ensemble models such as Random Forest (RF) and XGBoost have demonstrated superior performance in geologically variable settings, enabling robust classification [2, 9, 10]. These algorithms have been applied across various contexts, including shale–quartzite discrimination, identification of lithological units in banded iron formations, and mapping of volcanic–fluvial lithofacies [11, 12]. Recent studies have also highlighted the utility of remote sensing and well log data in lithological prediction [13, 14].

Corresponding author: mcotrinat@unitru.edu.pe (M. Cotrina)

Additional methods, such as deep neural networks (CNN, MLP), probabilistic classifiers (Naive Bayes, discriminant analysis), and hybrid frameworks integrating clustering, dimensionality reduction (Kernel PCA), and ensemble learning, have been explored to handle high-dimensional and noisy geological data [15–19]. These approaches have been successfully applied to well logs [20, 21], sedimentary formations, offshore wells [22], and real-time drilling scenarios [23].

Several authors have explicitly demonstrated that incorporating XYZ coordinates into MWD or exploration datasets significantly improves lithological discrimination and produces more geologically coherent 3D domains, especially when using ensemble tree models or neural networks [12, 24, 25]. Other studies have employed spatial variables in 3D geological prediction tasks such as mineral prospectivity or subsurface domain mapping, where spatial predictors are standard inputs for constructing volumetric models that preserve geological continuity [26, 27]. Consequently, spatial predictors have become standard components of contemporary lithological modeling workflows [28]. However, the integration of spatial data with uncertainty quantification remains limited: while many studies report confidence levels, they often fail to relate zones of high uncertainty to geological controls, reducing their practical value for targeting and domain delineation [12, 29].

Despite increasing interest in uncertainty analysis, machine learning studies on lithology still prioritize global classification metrics (accuracy, precision, recall, F1), often neglecting the spatial distribution of predictive reliability within the 3D model domain [30]. This is critical because drilling data are spatially autocorrelated, and contacts are gradational; thus, a model may achieve high overall accuracy yet perform poorly in transitional or sparsely sampled zones where geological risk is concentrated [31]. Pirot et al., Li et al., and Nie et al. [1, 32, 33] have begun operationalizing uncertainty from class probabilities, most commonly through Shannon entropy or joint entropy, to identify low-confidence zones and quantify prediction dispersion in lithological or domain models. Entropy-based approaches are appealing because they are model-agnostic, straightforward to compute from multiclass probabilities, and widely used in geological uncertainty assessment. However, these outputs often remain purely statistical: high-entropy zones are rarely interpreted in terms of geological controls or sampling limitations, limiting their usefulness for domain delineation and exploration targeting [1, 34].

To address these gaps, this study presents a systematic comparison of six supervised machine learning algorithms—Random Forest, XGBoost, SVM, KNN, Naive Bayes, and ANN-MLP—applied to the lithological classification of a structurally complex gold deposit in northern Peru. The models were trained using both geochemical data (gold grade) and spatial coordinates (X, Y, Z).

Crucially, we incorporate an entropy-based uncertainty analysis to evaluate prediction confidence and interpret ambiguous geological zones.

This study contributes to lithological modeling by combining uncertainty quantification and ML, offering a robust, interpretable framework for application in complex mineral systems.

2. Methodology

2.1. Geological setting and methodology

This study was conducted at a gold deposit located in the Cerro Pampa Larco area in the northern region of Peru [35]. The deposit lies at an average elevation of 3,500 meters above sea level, approximately 11.5 km northeast of the town of Angasmarca, in the district of Angasmarca, province of Santiago de Chuco, La Libertad region, and 164 km east of the city of Trujillo.

The deposit is situated at the northern end of Peru's northern gold belt, a metallogenic province characterized by disseminated gold deposits primarily hosted in Lower Cretaceous sedimentary rocks. The dominant geological unit in the area is the Chimú Formation, composed of medium- to coarse-grained quartzites and sandstones, interbedded with siltstones, shales, and thin coal layers [36].

These sedimentary units exhibit high secondary porosity and permeability due to intense structural deformation, making them favorable hosts for the circulation of hydrothermal fluids and the subsequent precipitation of gold mineralization. The mineralization is primarily of the disseminated type and is associated with iron oxides and sulfates (goethite, jarosite), forming under highly oxidized conditions and extending to depths greater than 250 meters [37].

Structural and stratigraphic controls play a critical role in gold concentration, particularly in areas where shale and siltstone layers act as traps for mineralizing fluids. These conditions have resulted in brecciated and fractured zones enriched in free gold, quartz, limonite, and clays.

The lithology of the deposit comprises three main units, clearly distinguished along the SW–NE geological section (Figure 1): Chimú Formation – Member B (chim-MB): composed of sandstones and quartzites with coal intercalations in the upper portion. These rocks possess physical properties favorable for hosting gold mineralization. Chimú Formation – Member A (chim-MA): consists mainly of medium-grained quartzites with less development of interbedded layers. This unit is also considered favorable for mineralization. Chicama Formation: the underlying unit beneath the Chimú Formation, composed of dark shales and siltstones interbedded with sandstones and quartz lenses. While generally less favorable for economic mineralization, it may contain mineralization near contact with intrusive bodies.

The presence of dacitic porphyries and potassic-ferric (K-F) porphyries, associated with the Late Miocene, further complements the geological framework of the area, suggesting potential involvement in hydrothermal alteration processes related to gold mineralization.

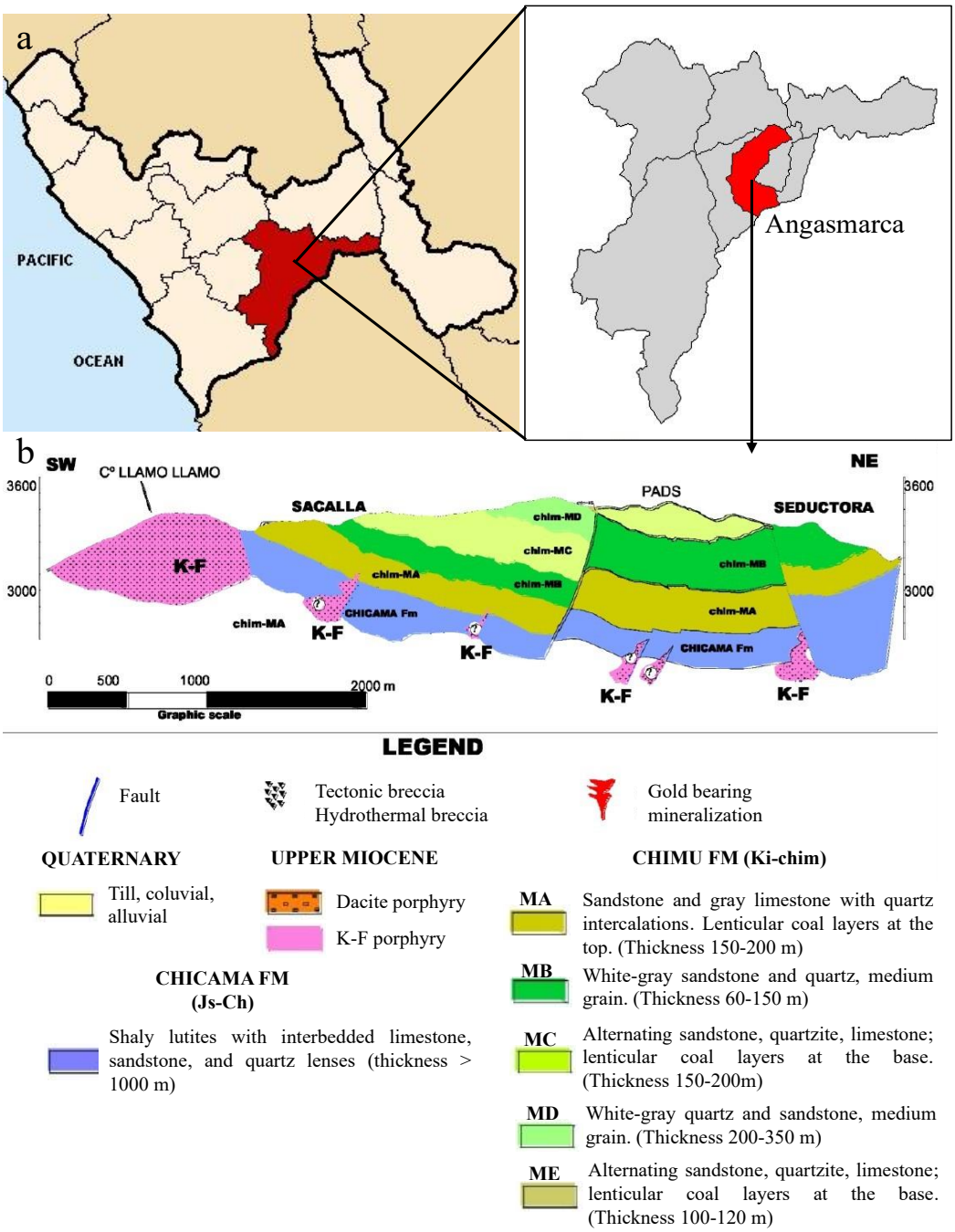

Figure 1. (a) Regional location map of the study area in northern Peru. (b) SW-NE geological cross-section of the study area within the gold deposit in Peru [35]

2.2. Database description

A general workflow diagram summarizing the methodology employed in this study is shown in Figure 2. It outlines the process from data collection to result evaluation, including exploratory analysis, data preprocessing, application of classification models, and uncertainty analysis.

This study is based on data collected during the exploration phase of a gold deposit located in the northern region of Peru. The dataset comprises 2,129 composites derived from 140 diamond drill holes spatially distributed across the study area. Each composite corresponds to an 8-meter interval, which matches the planned bench height for future open-pit operations. This compositing length was selected not only for operational consistency but also because it reflects the average thickness of the main lithological and mineralized zones identified during geological logging. This alignment reduces boundary fragmentation, enhances the spatial coherence of the predictive models, and facilitates direct integration into mine planning workflows [38].

The dataset integrates both spatial and geological variables. The spatial variables include East, North, and Elevation coordinates for each interval (see Figure 3). The geological variable corresponds to lithology, categorized into three main types according to the intercepted stratigraphic unit:

- Type 80: Member B of the Chimú Formation (sandstones and quartzites with coal layers)
- Type 85: Member A of the Chimú Formation (medium-grained quartzites)
- Type 90: Chicama Formation (shales, siltstones, and sandstones with quartz lenses)

Additionally, a quantitative geochemical variable is included: gold grade (Au), reported in grams per tonne (g/t), obtained through laboratory analysis of core samples from the drill holes.

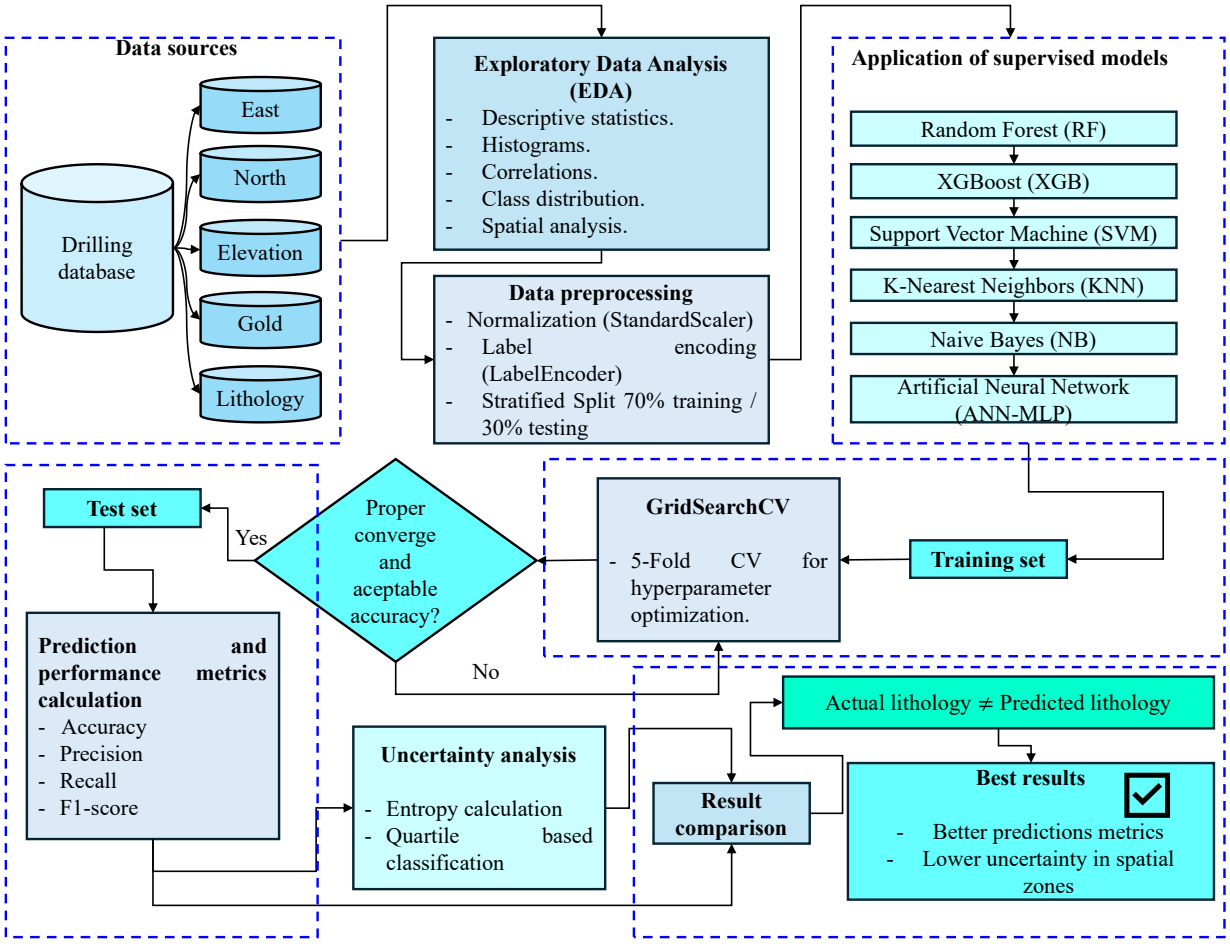

Figure 2. Methodological workflow applied for lithological classification

The variables selected for training the models include three spatial coordinates (East, North, and Elevation) and one geochemical variable (gold content in g/t). These variables were chosen due to their proven geological relevance in the characterization of lithostratigraphic units. The

spatial coordinates help capture patterns of continuity and structural relationships within the 3D domain of the deposit, such as lithological contacts, faults, and mineralized zones associated with intrusive bodies. Conversely, gold content serves as a geochemical proxy that reflects mineralization and hydrothermal alteration processes, which tend to be spatially correlated with certain lithological types.

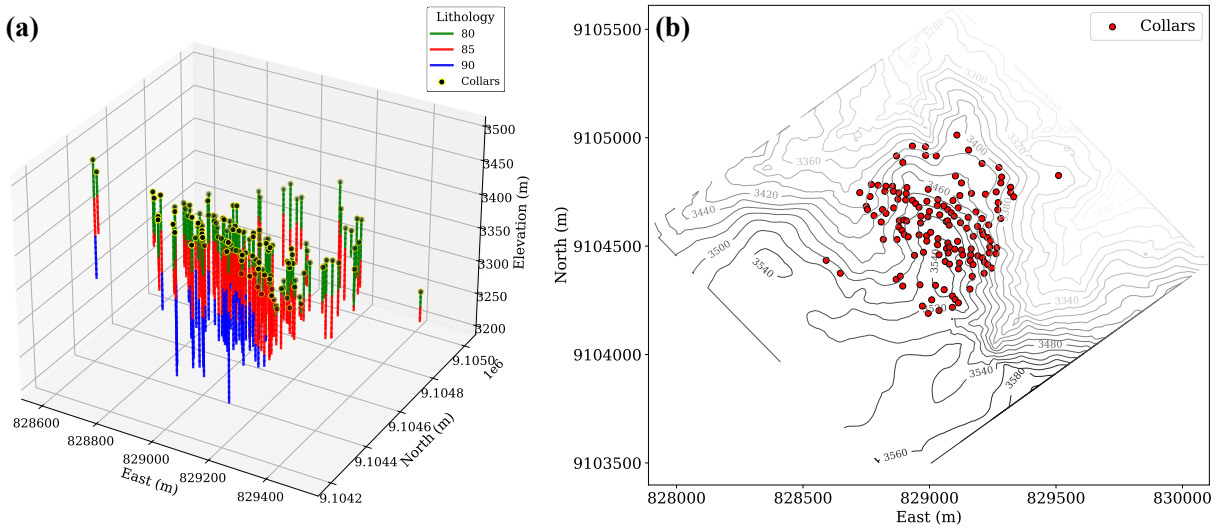

Figure 3. Spatial distribution of diamond drill holes in the study area. (a) Three-dimensional visualization with lithological coding. (b) Collar map overlaid on topographic contours

2.3. Data processing

Prior to training the classification models, data preprocessing procedures were applied to ensure the consistency and quality of the input variables. This stage involved three key steps: normalization of continuous variables, encoding of categorical labels, and stratified splitting of the dataset.

Continuous numerical variables were standardized to zero mean and unit variance using the StandardScaler implementation in scikit-learn [39]. This technique transforms each individual value x_i of variable X based on its deviation from the mean, ensuring that all variables have a mean of zero and unit variance. See Equation (1):

$$x_i^{norm} = \frac{x_i - \mu X}{\sigma X} \quad (1)$$

where x_i^{norm} is the normalized value, x_i is the original value, μX is the mean of variable X and σX is the standard deviation of X.

The target variable for the model is lithology, a nominal categorical variable with three distinct classes: Type 80, Type 85, and Type 90. Label encoding was applied to convert categorical lithology classes into a numerical format interpretable by classifiers [39]. This process assigns an integer value to each lithological class without introducing artificial ordinal relationships.

To ensure generalization and prevent information leakage, the dataset was split at the drillhole level, not the sample level. This means that all samples from a given drillhole were assigned entirely to either the training or test set, avoiding cross-contamination between spatially autocorrelated observations. A stratified approach was applied at the drillhole level to preserve the global class distribution of lithologies across both subsets. This strategy reflects the spatial structure of the data and provides a more realistic evaluation of model generalization in unseen regions [39].

2.4. Supervised classification models

This study employs six supervised machine learning algorithms—Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes (NB), and Artificial Neural Network with Multilayer Perceptron (ANN-MLP)—to perform lithological classification. The selection is based on three criteria: (i) documented success in geoscientific classification problems, (ii) capacity to integrate multivariate data, including spatial (X, Y, Z) and geochemical attributes, and (iii) diversity in methodological paradigms. This diversity allows for a comprehensive comparison of predictive accuracy, uncertainty quantification, and

interpretability in complex geological environments [12, 17, 34].

2.4.1. Random Forest

Random Forest, a widely used ensemble learning technique, has demonstrated high reliability in geological settings due to its ability to capture nonlinear relationships and accommodate missing or noisy data. It constructs multiple decorrelated decision trees using bootstrapped subsets and averages their outputs, which reduces overfitting and enhances robustness [40]. RF has been frequently applied to lithological mapping, ore type classification, and mineral prospectivity modeling [12, 27, 41–44], making it a strong baseline model in this study.

2.4.2. Extreme Gradient Boosting

XGBoost is a gradient-boosted tree algorithm that sequentially builds models by minimizing residual errors of prior trees. It includes regularization, parallelization, and optimized handling of missing data, making it particularly effective in high-dimensional classification tasks. Its application in subsurface modeling has grown rapidly in recent years, especially in the classification of stratigraphic units and alteration zones [45–47], justifying its inclusion as a high-performance ensemble alternative.

2.4.3. Support Vector Machine

SVM constructs hyperplanes in high-dimensional spaces to maximize class separation and is particularly effective for problems where classes are not linearly separable. Its robustness against overfitting and good performance with limited samples make it suitable for geological classification with moderate datasets [12, 48, 49]. Kernel functions further enhance its flexibility in modeling complex lithological boundaries.

2.4.4. K-Nearest Neighbors

KNN is a non-parametric, instance-based learner that classifies samples based on their proximity to labeled neighbors. Although sensitive to noise and dimensionality, its simplicity makes it a useful benchmark, especially when spatial continuity is assumed. In mineral exploration, KNN has been used to delineate ore zones and lithological classes where neighborhood effects are geologically meaningful [50–52].

2.4.5. Naive Bayes

NB is a probabilistic classifier grounded in Bayes' theorem, assuming conditional independence among features. Despite this simplification, it has proven effective in multiclass classification tasks, particularly when datasets are sparse or imbalanced. In geoscientific contexts, NB has been used for early-stage lithological classification due to its rapid training and interpretability [53–55].

2.4.6. Artificial Neural Network-Multilayer Perceptron

ANN-MLP is a deep learning architecture that models complex nonlinear relationships through multiple hidden layers. Its capacity to generalize over noisy, high-dimensional data has made it increasingly popular in geoscience applications, including lithology prediction from well logs and drillhole datasets [56–59]. The flexibility of the MLP framework also supports the integration of diverse input features, including spatial coordinates and geochemical signatures.

Together, these models span a broad spectrum of machine learning paradigms—tree-based, kernel-based, distance-based, probabilistic, and deep learning—ensuring a representative and comparative evaluation of classification performance and uncertainty estimation in lithologically complex terrains.

2.5. Validation and performance metrics

Rigorous evaluation of model performance is essential to ensure the robustness and generalizability of lithological predictions in geologically complex contexts.

2.5.1. Spatial aware date splitting

To prevent data leakage caused by spatial autocorrelation between samples from the same drillhole, the dataset was partitioned at the drillhole level rather than at the individual sample level. This approach ensures that all samples from a given drillhole are either entirely in the training set or the test set, eliminating the possibility of information from one portion of the drillhole influencing the model's performance on another [60, 61]. This methodology preserves the independence of validation data and provides a more realistic assessment of model generalization in operational scenarios.

2.5.2. Stratified K-Fold cross-validation

To reduce model variance and avoid overfitting, a 5-fold cross-validation scheme ($K = 5$) was employed. In each iteration, the dataset was divided into five approximately equal groups, with four folds used for training and one for validation. Stratification ensured that the distribution of lithological classes was preserved across all folds. This method yields a robust estimate of the model's average performance across the full dataset [62].

2.5.3. Data quality and sampling control

The original dataset comprises samples obtained from diamond drilling conducted across a structurally complex gold deposit in northern Peru. Drillholes were logged by certified geologists following standardized protocols for lithological classification and gold grade quantification. Prior to model development, the dataset was screened for inconsistencies, outliers, and missing values [12, 63]. Quality control procedures included visual validation of geochemical anomalies, spatial validation of coordinates, and cross-referencing against validated geological models. This ensured a reliable foundation for model training and evaluation.

2.5.4. Confusion matrix and ROC curve

The confusion matrix was used to analyze the detailed performance of each model with respect to individual lithological classes. This matrix summarizes correct versus incorrect predictions, allowing the identification of classification bias, confusion between similar lithologies, and systematic errors. In the multiclass case, the confusion matrix takes the form of a square matrix of order C, where C is the number of lithological classes. Each cell (i, j) indicates the number of actual observations from class that were classified as class j [64, 65].

In addition, ROC curves were generated for each model using a one-vs-rest approach to adapt the ROC curve to the multiclass setting. The ROC curve evaluates the trade-off between the true positive rate (TPR) and the false positive rate (FPR). See Equation (2):

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (2)$$

where TP refers to true positives, FN to false negatives, FP to false positives, and TN to true negatives.

Each ROC curve illustrates the model's ability to distinguish one class from the others. Additionally, the area under the curve (AUC) was computed as an integral measure of model performance; an AUC value close to 1 indicates high discriminative power [11].

2.5.5. Performance metrics

Based on the confusion matrix, four standard metrics were calculated: accuracy, precision, recall, and F1-score [11, 12].

Accuracy: represents the proportion of correct predictions out of the total number of observations. It serves as a general measure of overall correctness. See Equation (3):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Precision: indicates how many of the samples predicted as positive are actually positive. See Equation (4):

$$Precisión = \frac{TP}{TP + FP} \quad (4)$$

Recall: measures the proportion of actual positives that were correctly identified. It is useful for evaluating how many relevant cases were retrieved by the model. See Equation (5):

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

F1-Score: combines precision and recall into a single metric, especially useful when dealing with imbalanced datasets. See Equation (6):

$$F1 - Score = 2 * \frac{Precisión * Recall}{Precisión + Recall} \quad (6)$$

2.6. Uncertainty evaluation

To assess the reliability of lithological predictions, a predictive uncertainty analysis was incorporated. Entropy, in this context, refers to the degree of uncertainty in the model's classification decision. It measures how evenly the model distributes probability across the possible lithological classes. When a model is confident, it assigns a high probability to one class, resulting in low entropy. Conversely, if it is uncertain, the probabilities are more evenly distributed, leading to high entropy. This makes entropy a useful tool for identifying predictions where the model lacks certainty, such as lithological boundaries or areas with sparse data [66–69]. Mathematically, for an observation with C classes and associated probabilities $p_1, p_2, \ldots, p_c$, the entropy is defined in Equation (7):

$$H(x) = -\sum_{i=1}^{C} p_i log_2(p_i) \quad (7)$$

where p_i represents the probability assigned to class i, and $H(x)$ is the level of predictive uncertainty for that observation. Based on this metric, the samples were grouped into four uncertainty levels (very low, low, medium, and high) using the 25th, 50th, and 75th percentiles of the entropy distribution for each model. Subsequently, the performance of the supervised classification models was evaluated within each uncertainty level using standard metrics such as accuracy and F1-score. This approach allows for the identification of the stability and robustness of each algorithm under varying levels of geological ambiguity.

3. Results and discussion

3.1. Exploratory data analysis

The exploratory analysis enabled the characterization of spatial, geochemical, and lithological variables within the dataset used for classification. As shown in Table 1, the spatial variables East, North, and Elevation exhibit a homogeneous distribution. Elevation shows significant variation, which is consistent with the morphostructural context of the Andes Mountain Range, where lithological transitions are often controlled by folds, faults, and stratigraphic traps [28].

Table 1. Descriptive statistics of spatial and geochemical variables in the database

Statistics	East (m)	North (m)	Elevation (m)	Gold (g/t)	Lithology
Count			2,129.00		
Mean	829,036.89	9,104,555.57	3,328.54	0.234	-
Std. Dev.[a]	150.94	184.37	56.30	0.317	-
Minimum	828,591.37	9,104,189.68	3,207.04	0.003	80.00
Median	829056.95	9104542.85	3,321.57	0.138	85.00
Maximum	829510.01	9105012.42	3493.02	4.104	90.00
Variance	22,783.39	33,992.28	3,170.02	0.10	-
Range	918.63	822.74	285.98	4.10	10.00
Coef. Var (%)[b]	0.018	0.002	1.692	135.12	-
Mad[c]	121.99	150.27	45.47	0.19	-

a) Std.Dev: standard deviation; b) Coef. Var: coefficient of variation; c) MAD: median absolute deviation

The geochemical variable, gold grade (Au), exhibits a strongly right-skewed distribution, with a mean value of 0.234 g/t and a maximum of 4.10 g/t (Figure 4). This pattern indicates highly heterogeneous mineralization, characteristic of disseminated or structurally controlled deposits [2]. The high coefficient of variation (CV = 135.12%) and the presence of outliers support the use of machine learning algorithms capable of capturing complex and nonlinear relationships within the data.

Although the dataset shows a moderate imbalance among lithological classes, with class 85 being the most prevalent (51.2%), followed by class 80 (30.3%) and class 90 (18.4%), measures were taken to mitigate potential biases during model training. A stratified partitioning scheme was applied both in the train-test split and during cross-validation, ensuring proportional class representation across all subsets. This strategy improves the representation of minority classes and reduces the risk of overfitting toward the dominant class. This issue has been addressed by Shebl et al. [8], who recommend the use of techniques such as stratified sampling or alternative evaluation metrics to prevent bias.

3.2. Data preprocessing and hyperparameters tuning

Before training the supervised classification models, a rigorous data preparation process was conducted. The dataset was divided into two subsets: 70% for training and 30% for testing, ensuring a stratified distribution of the target variable (lithology) to maintain class balance in both groups. The descriptive statistics for each partition (Table 2) show similar distributions in terms of mean, standard deviation, and value range for the spatial and geochemical variables.

The resulting class distribution after dataset partitioning is shown in Figure 5. An appropriate balance among class proportions is observed in both subsets, preserving the statistical representativeness of the test set: lithology class 85 accounts for more than 50% of the samples, followed by class 80, and finally class 90.

Table 2. Descriptive statistics of spatial and geochemical variables in the training and test sets

Statistics	Training				Test			
	East (m)	North (m)	Elevation (m)	Gold (g/t)	East (m)	North (m)	Elevation (m)	Gold (g/t)
Mean	829,037.55	9,104,554.30	3,328.42	0.23	829,035.34	9,104,558.54	3,328.82	0.24
Std. Dev.	150.64	184.14	56.25	0.32	151.76	185.02	56.46	0.30
Minimum	828,591.37	9,104,189.68	3,208.96	0.003	828,591.37	9,104,189.68	3,207.04	0.003
Maximum	829,510.01	9,105,012.42	3,493.02	4.10	829,510.01	9,105,012.42	3,488.76	3.11

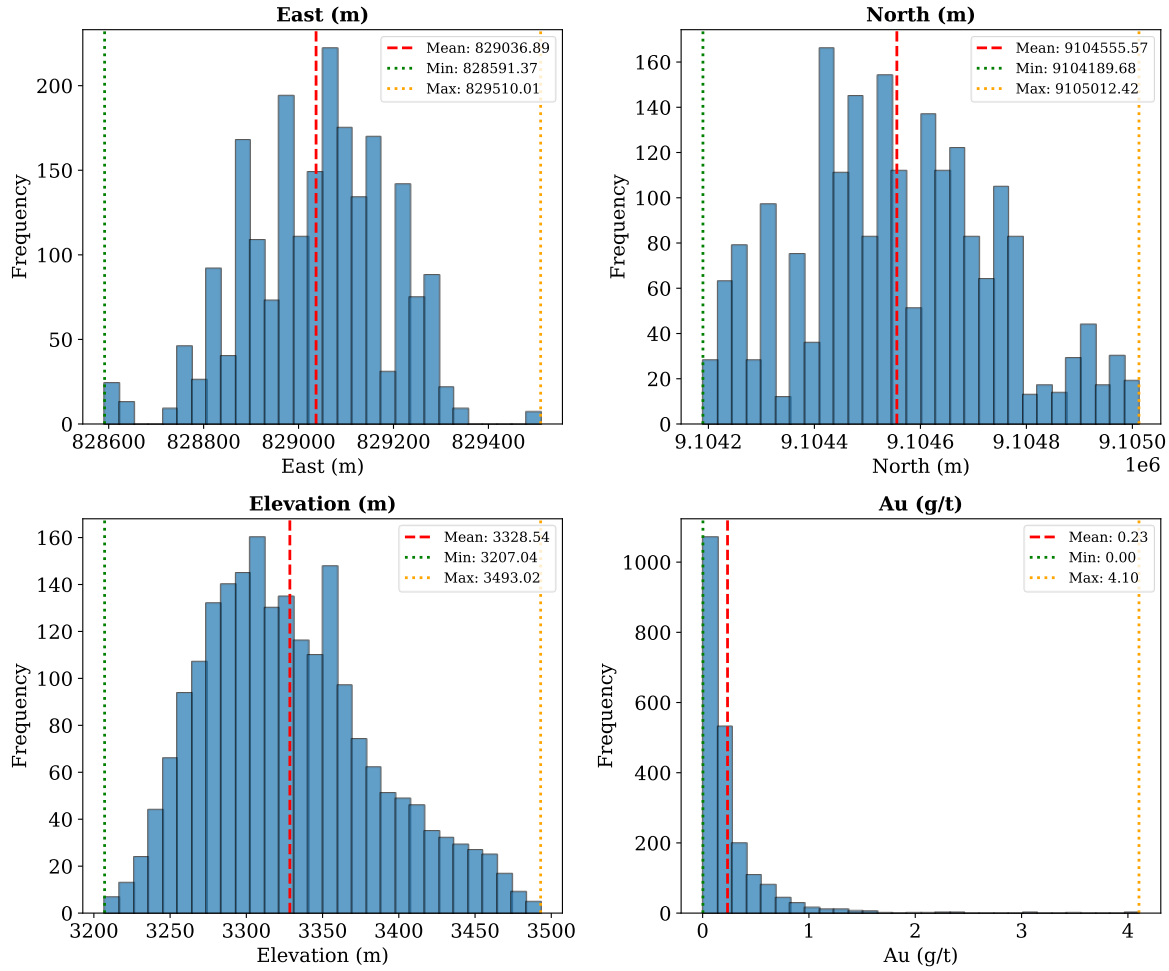

Figure 4. Frequency distribution of numerical variables: spatial coordinates and gold content

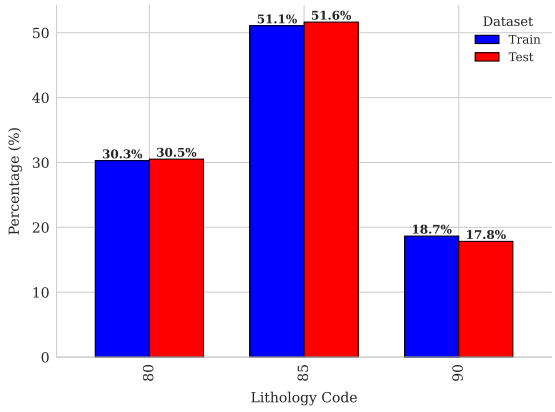

Figure 5. Proportional distribution of lithological classes in the training and test sets

The categorical lithology variable was encoded using the LabelEncoder method, assigning consecutive integer values (0, 1, 2) to the original codes (80, 85, 90). Additionally, the numerical variables were normalized using standard scaling (StandardScaler), adjusting the data to zero mean and unit standard deviation. This transformation is critical for scale-sensitive algorithms such as SVM and neural networks [17].

The selected hyperparameter ranges were defined based on prior studies in geoscientific machine learning applications and preliminary tuning experiments. For tree-based models (RF, XGBoost), parameters such as the number of estimators, maximum tree depth, and minimum

samples per split were varied following ranges commonly used in lithological classification tasks [12, 43]. For SVM, kernel types and regularization parameters were chosen to explore both linear and nonlinear class separability, in line with practices reported by Dong et al. [48] and Zhang et al. [49]. For ANN-MLP, configurations were inspired by architectures successfully applied to drillhole data and well logs [57], while regularization parameters were adjusted to prevent overfitting. For KNN and NB, simplicity and interpretability were prioritized, selecting ranges that balance computational efficiency and predictive stability. This strategy ensured that the search space was sufficiently broad to identify optimal configurations while remaining computationally feasible [70, 71].

For the specific case of the ANN-MLP model, different architecture configurations and regularization parameters were evaluated to minimize overfitting risk. The optimization process included systematic testing by varying the number of hidden layers (1 to 3), the number of neurons per layer (10 to 100), the activation function (ReLU and tanh), learning rate, and batch size. Additionally, L2 regularization was applied (with α values between 0.0001 and 0.01), and a dropout layer (range 0.1 to 0.3) was incorporated during training to reduce dependency on individual neurons. The final selection criterion was the average F1-score obtained during stratified cross-validation. The evolution of the loss function on both training and validation sets was also monitored to confirm the absence of overfitting. The optimal hyperparameters are summarized in Table 3.

Table 3. Hyperparameter search ranges and optimal values for supervised classification models

ML model	Hyperparameter	Description	Search range	Best value
RF	N° estimators	Number of trees in the forest	[50, 100, 150, 200, 250, 300]	100
	Max depth	Maximum depth of each tree	[5, 10, 15, 20, 25]	20
	Min samples split	Minimum number of samples to split a node	[2, 5, 10, 15, 20, 25]	5
	Min samples leaf	Minimum number of samples at a leaf node	[2, 4, 6, 8, 10, 12]	2
XGBoost	N° estimators	Number of boosting rounds	[50, 100, 150, 200, 300]	300
	Max depth	Maximum tree depth	[3, 5, 7, 10]	3
	Learning rate	Step size shrinkage used in updates	[0.01, 0.05, 0.1, 0.2]	0.1
	Subsample	Fraction of samples used for training each tree	[0.6, 0.8, 1.0]	0.8
	Colsample bytree	Fraction of features used per tree	[0.6, 0.8, 1.0]	1.0
SVM	C	Regularization parameter	[0.1, 1, 10, 100]	100
	Kernel	Kernel type used in the algorithm	[“linear”, “rbf”, “poly”]	“rbf”
	Gamma	Kernel coefficient	[“scale”, “auto”]	“auto”
	Degree	Degree of polunomial kernel	[2, 3, 4]	2
KNN	N° neighbors	Number of nearest neighbors	[3, 5, 7, 9, 11]	9
	Weights	Weight function used in prediction	[“uniform”, “distance”]	“distance”
	Metric	Distance metric used	[“euclidean”, “manhattan”, “minkowski”]	“manhattan”
	P	Power parameter for the minkowski metric	[1, 2]	1
NB	Var smoothinga	Portion of variance added for numerical stability	7 valores entre [1e-9 y 1e-3]	1e-9
ANN-MLP	Hidden layer sizes	Number of neurons in hidden layers	[(50), (100), (50,50), (100,50)]	100,50
	Activation	Activation function for hidden layers	[“relu”, “tanh”, “logistic”]	“tanh”
	Solver	Optimization algorithm used	[“adam", “sgd”]	“adam”
	Alpha	L2 regularization term (weight decay)	[1e-5, 1e-4, 1e-3]	0.001
	Learning rate	Learning rate schedule	[“constant”, “adaptive”]	“constant”

3.3. Machine learning models for lithological classification

Table 4 presents the 5-fold cross-validation results for the six machine learning models evaluated. The ANN-MLP model achieved the best overall performance during the training phase, with an average accuracy of 0.938 and an F1-score of 0.932, indicating strong predictive power and stability in lithological classification. XGBoost and Random Forest also performed well, with scores above 0.91, highlighting their robustness as ensemble models. In contrast, Naive Bayes showed lower performance, with an F1-score of only 0.613, confirming its limitations in geological contexts where variables exhibit complex correlations.

Table 4. Cross-validation results: average metrics and standard deviation on the training set for the machine learning models

Machine learning model	Accuracy (cv mean ±std)	Precision (cv mean ±std)	Recall (cv mean ±std)	F1-score (cv mean ±std)	Accuracy (cv mean ±std)
RF	0.909 ± 0.002	0.911 ± 0.003	0.909 ± 0.002	0.909 ± 0.002	0.909 ± 0.002
XGBoost	0.919 ± 0.015	0.919 ± 0.016	0.919 ± 0.015	0.919 ± 0.016	0.919 ± 0.015
SVM	0.908 ± 0.017	0.908 ± 0.017	0.908 ± 0.017	0.908 ± 0.017	0.908 ± 0.017
KNN	0.875 ± 0.022	0.877 ± 0.021	0.875 ± 0.022	0.875 ± 0.022	0.875 ± 0.022
NB	0.633 ± 0.021	0.657 ± 0.021	0.633 ± 0.021	0.613 ± 0.029	0.633 ± 0.021
ANN-MLP	0.932 ± 0.014	0.933 ± 0.014	0.932 ± 0.014	0.932 ± 0.014	0.932 ± 0.014

The results obtained by each model on the test set are summarized below. The ANN-MLP model achieved the best overall performance, with accuracy, precision, recall, and F1-score all equal to 0.945, validating its ability to generalize effectively to unseen data. It was closely followed by XGBoost, which showed nearly identical metrics (0.939 across all measures), establishing itself as a highly effective alternative. Random Forest also delivered strong results, achieving 0.922 accuracy, while SVM and KNN maintained acceptable performance levels. Naive Bayes recorded the lowest performance, reflecting its sensitivity to non-Gaussian distributions and the lack of independence among predictive variables (see Table 5).

Table 5. Performance metrics of machine learning models on the test set

Machine learning model	Accuracy (test)	Precision (test)	Recall (test)	F1-score (test)	Accuracy (test)
RF	0.922	0.924	0.922	0.921	0.922
XGBoost	0.939	0.940	0.939	0.939	0.939
SVM	0.908	0.908	0.908	0.908	0.908
KNN	0.886	0.900	0.886	0.885	0.886
NB	0.613	0.636	0.613	0.597	0.613
ANN-MLP	0.945	0.946	0.945	0.945	0.945

The superior performance of ANN-MLP and XGBoost can be attributed to their architectural strengths and suitability for the data structure. XGBoost, through its boosting strategy and regularization techniques, efficiently handles multicollinearity, outliers, and class imbalance characteristics often present in lithological and geochemical datasets [12]. It constructs additive trees sequentially, allowing it to model complex nonlinear relationships and refine boundary zones between lithologies [72]. Meanwhile, the ANN-MLP architecture with multiple hidden layers and nonlinear activation functions effectively captures high-dimensional interactions between spatial coordinates and geochemical features [57]. This capacity to learn abstract representations makes it particularly well suited for datasets with overlapping class boundaries and nonlinear geological patterns [58]. Both models exhibit high flexibility and generalization capacity, which explain their robustness in both global metrics and spatial coherence observed in 3D lithological mapping.

The confusion matrices provide insight into each model's performance in terms of correct and incorrect classifications across the three lithological classes (80, 85, and 90). ANN-MLP and XGBoost achieved the highest diagonal accuracy rates, indicating excellent differentiation capabilities between lithologies. In contrast, Naive Bayes showed the highest number of misclassifications, particularly confusing lithology 90 with 85. This suggests a limited ability to handle classes with spatial or compositional overlap (see Figure 6).

The multiclass ROC curves highlight the discriminative performance of each model in correctly classifying the different lithologies. The ANN-MLP model yielded AUC values close to 1 for all classes, with a micro-average AUC of 0.995, indicating an excellent balance between sensitivity and specificity. XGBoost and RF also produced outstanding ROC curves. In contrast, Naive Bayes exhibited flatter curves, especially for class 85 (AUC = 0.712), confirming its limitations in handling complex multiclass problems. These findings are consistent with [18], who reported similar results in geological exploration data (see Figure 7).

A three-dimensional visualization of lithological classification for each machine learning model shows that ANN-MLP, XGBoost, and RF produced clear and geologically coherent segmentations. These models successfully

captured lithological boundaries across the spatial domain. On the other hand, the Naive Bayes model yielded a more scattered and noisy classification, consistent with its lower quantitative performance (see Figure 8).

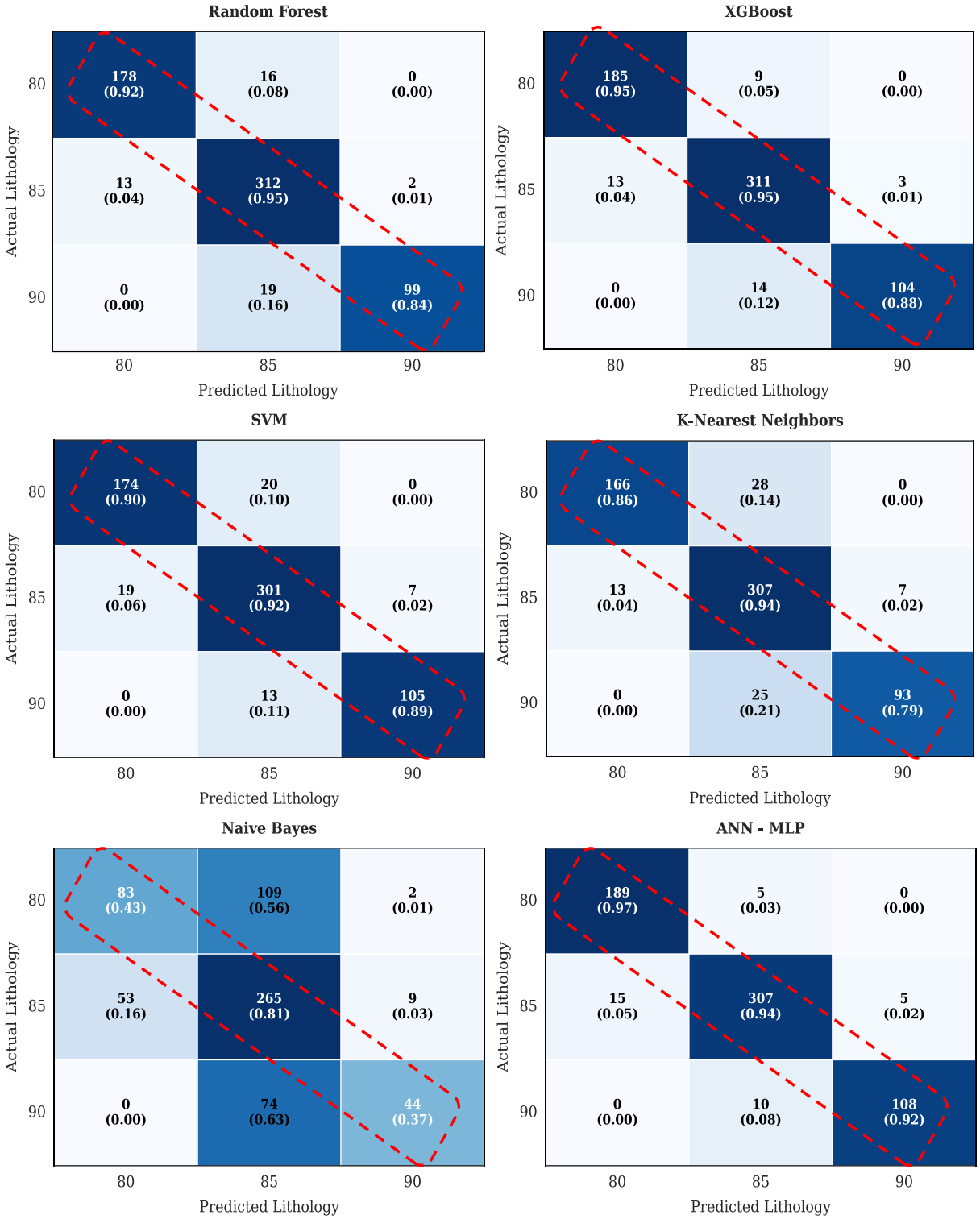

Figure 6. Confusion matrices for the machine learning classification models evaluated on the test set

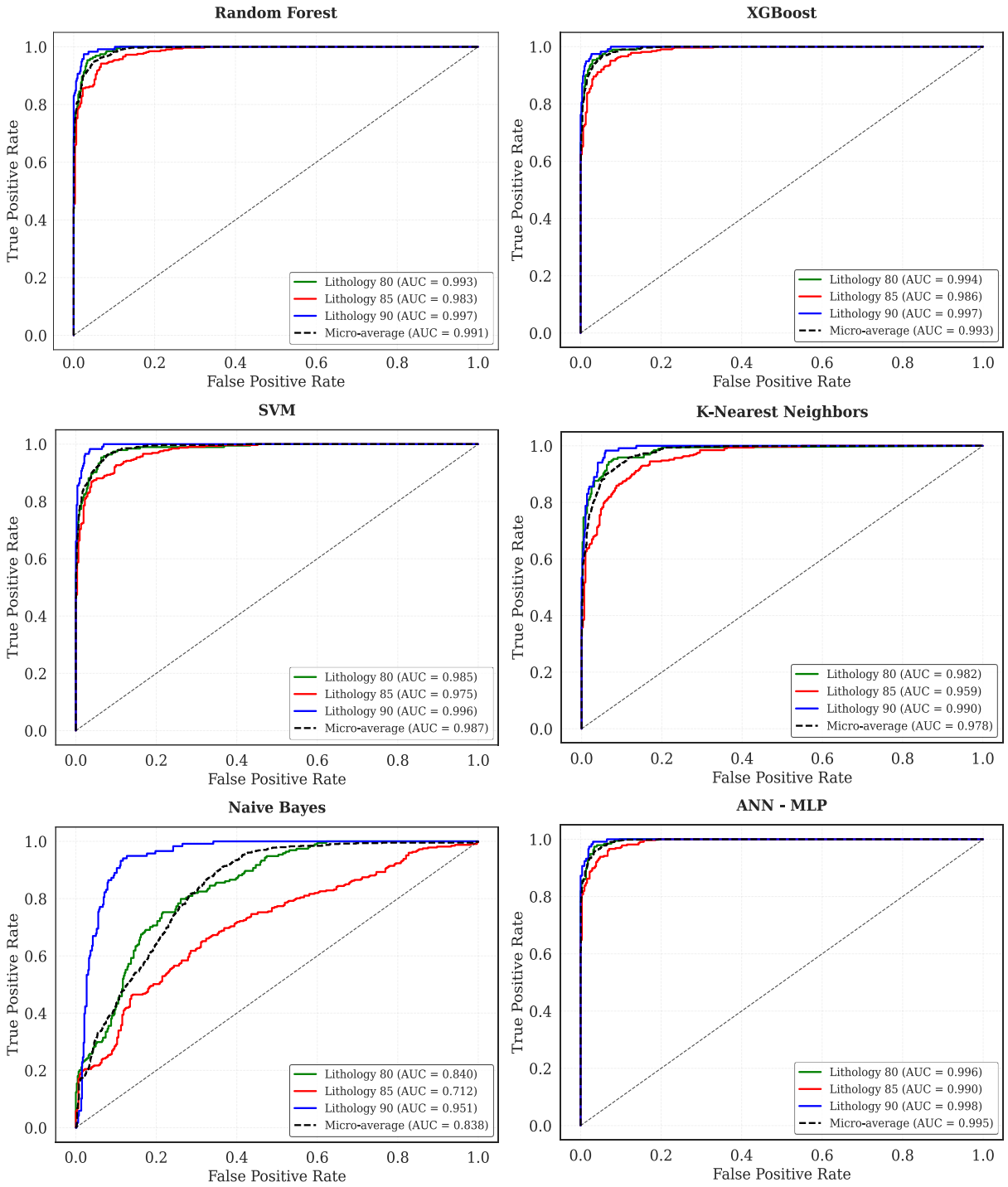

Figure 7. Multiclass ROC curves for the machine learning classification models evaluated on the test set

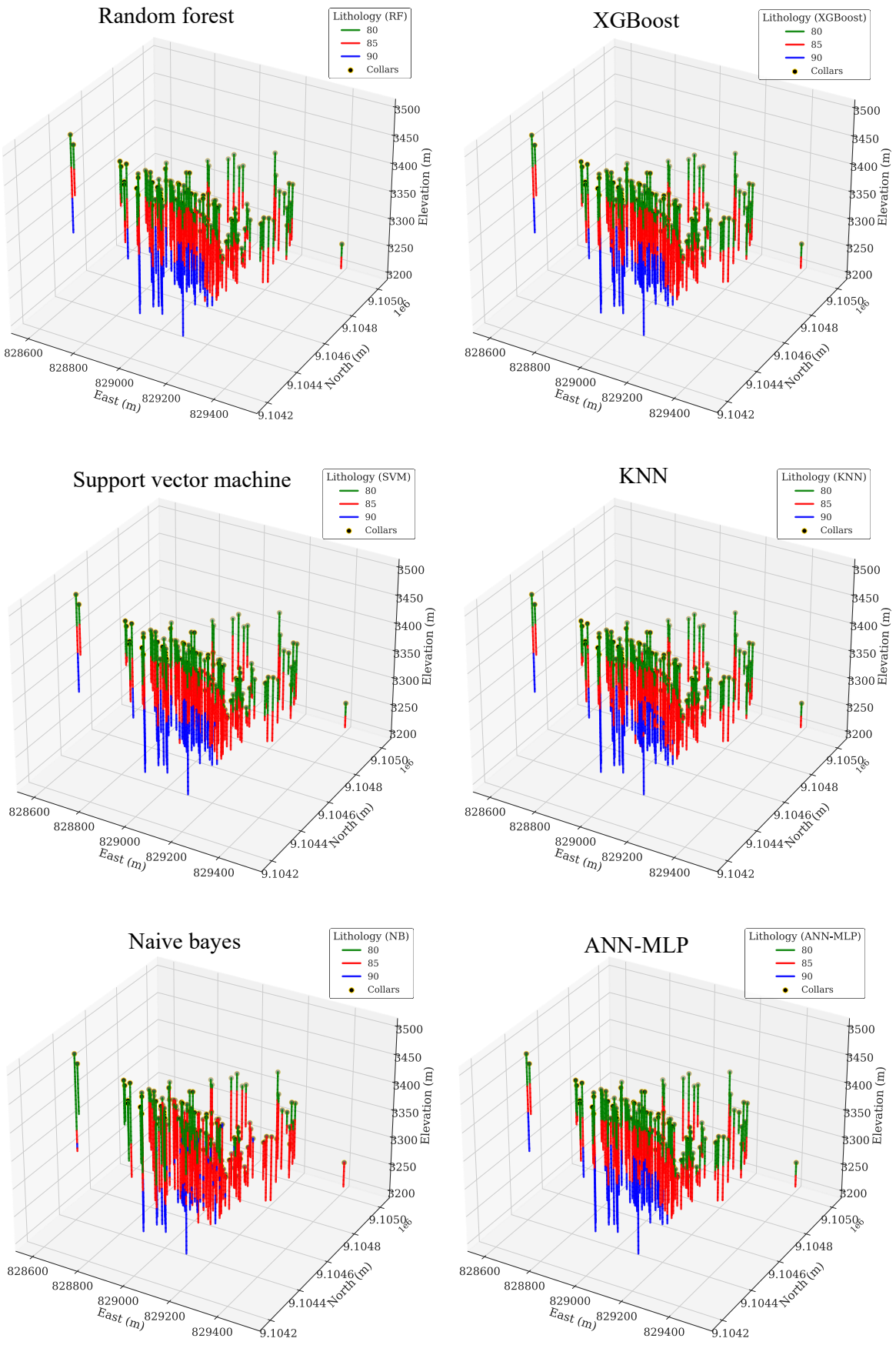

Figure 8. Three-dimensional visualization of lithological classification by machine learning model

3.4. Uncertainty analysis

The results in Table 6 clearly show that all supervised classification models exhibit a decreasing performance trend as predictive uncertainty, defined by entropy percentiles, increases. Models such as Random Forest, XGBoost, SVM, and ANN-MLP achieved 100% accuracy in the very low uncertainty category, suggesting high prediction confidence when entropy is minimal. As the uncertainty level increases, especially in the upper quartile, performance drops significantly. For instance, in Random Forest, the F1-score falls to 71.57% at the high uncertainty level. This behavior aligns with findings by Shashel (2022), who emphasizes that predictive entropy is an effective indicator of reliability in geological classification tasks.Haga clic o pulse aquí para escribir texto.

Table 6. Performance of supervised classification models by uncertainty level (defined by entropy percentiles)

ML model	Uncertainty level	N° of samples	Accuracy (%)	F1-score (%)
RF	Very low	160	100.00	100.00
	Low	160	99.38	99.37
	Médium	159	97.48	97.46
	High	160	71.88	71.57
XGBoost	Very low	160	100.00	100.00
	Low	160	100.00	100.00
	Médium	159	96.86	96.80
	High	160	78.75	78.63
SVM	Very low	160	100.00	100.00
	Low	159	99.38	99.37
	Médium	160	94.97	94.97
	High	160	68.75	68.84
KNN	Very low	190	100.00	100.00
	Low	130	97.69	97.67
	Médium	159	89.31	89.21
	High	160	66.88	66.94
NB	Very low	160	79.38	79.53
	Low	160	53.75	51.00
	Médium	159	51.57	50.67
	High	160	60.62	58.54
ANN-MLP	Very low	160	100.00	100.00
	Low	160	100.00	100.00
	Médium	159	99.37	99.37
	High	160	78.75	78.54

The distribution of entropy across the four defined uncertainty levels is shown in Figure 9. A clear separation between the ranges is observed, validating the use of percentiles for categorizing predictive uncertainty. Models such as XGBoost and ANN-MLP exhibit more compact and controlled distributions, even under high uncertainty levels, while Naive Bayes displays greater dispersion across all levels, reflecting its lower ability to distinguish between classes with confidence. This evidence is consistent with Fu et al. [16], who noted that tree-based models and neural networks tend to produce more reliable uncertainty distributions in multiclass classification contexts.

Quantitatively, the entropy distributions (Figure 9) demonstrate narrower interquartile ranges in ANN-MLP and XGBoost, indicating consistent classification confidence across samples. Their low entropy variance ($\sigma^2 < 0.2$) contrasts with the broader spread observed in Naive Bayes ($\sigma^2 > 0.6$), reflecting higher model indecisiveness. These differences are statistically significant ($p < 0.01$) based on Levene's test for equality of variance, reinforcing the superior stability of deep and ensemble models in managing geologically noisy datasets.

The three-dimensional visualization in Fig. 10 highlights specific areas within the deposit where the models exhibit higher uncertainty. Across all models, boundary zones and lithological transitions show increased entropy, which is expected in complex geological settings where overlapping units can confuse classification algorithms. In contrast, the core zones of each lithological unit display lower uncertainty. This finding supports the utility of spatial entropy analysis for enhancing geological interpretation, as proposed by Ibrahim et al. [17]. Notably, the Naive Bayes model performs poorly in this regard, showing high uncertainty even in geologically homogeneous areas.

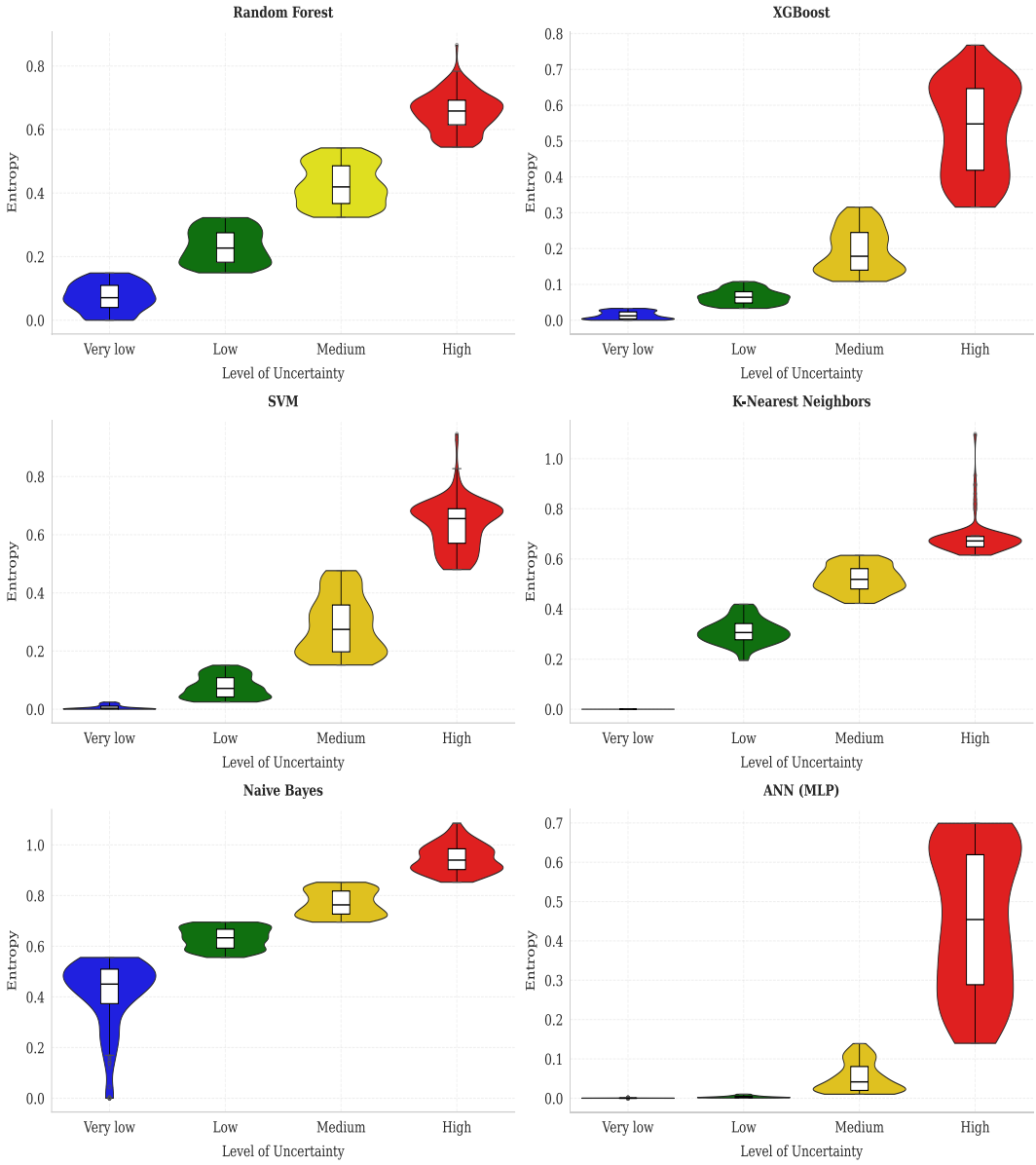

Figure 9. Distribution of predictive entropy by uncertainty level for each supervised classification model

The spatial patterns of high entropy observed in the 3D uncertainty maps (Figure 10) reveal valuable geological insights when interpreted within the structural and stratigraphic framework of the deposit (see Figure 1). These high-uncertainty zones are not randomly distributed; instead, they predominantly align with lithological boundaries and structural discontinuities. Notably, elevated entropy values are frequently concentrated along the contact zones between the quartz-rich sedimentary units of the Chicama Formation and the interbedded sandstone–limestone members of the Chimu Formation (chlm-MC to chlm-ME). These interfaces mark abrupt lithological changes and are commonly associated with structural complexity, including folding, faulting, and overprinting by hydrothermal processes. In particular, the areas surrounding the K-F porphyry intrusions and tectonic breccias exhibit significantly higher uncertainty levels. These features are interpreted as zones of structural and hydrothermal disturbance, where the original stratigraphy has been modified by the emplacement of magmatic bodies and associated fluid flow. The presence of high entropy in these regions likely reflects the difficulty of the models in classifying lithology under conditions of strong geochemical alteration, structural disruption, and ambiguous geospatial signatures. This is consistent with the observed spatial dispersion of entropy in models such as Naive Bayes, and the more localized high-entropy zones seen in ANN-MLP and XGBoost models, which are better equipped to manage non-linearities and spatial heterogeneity.

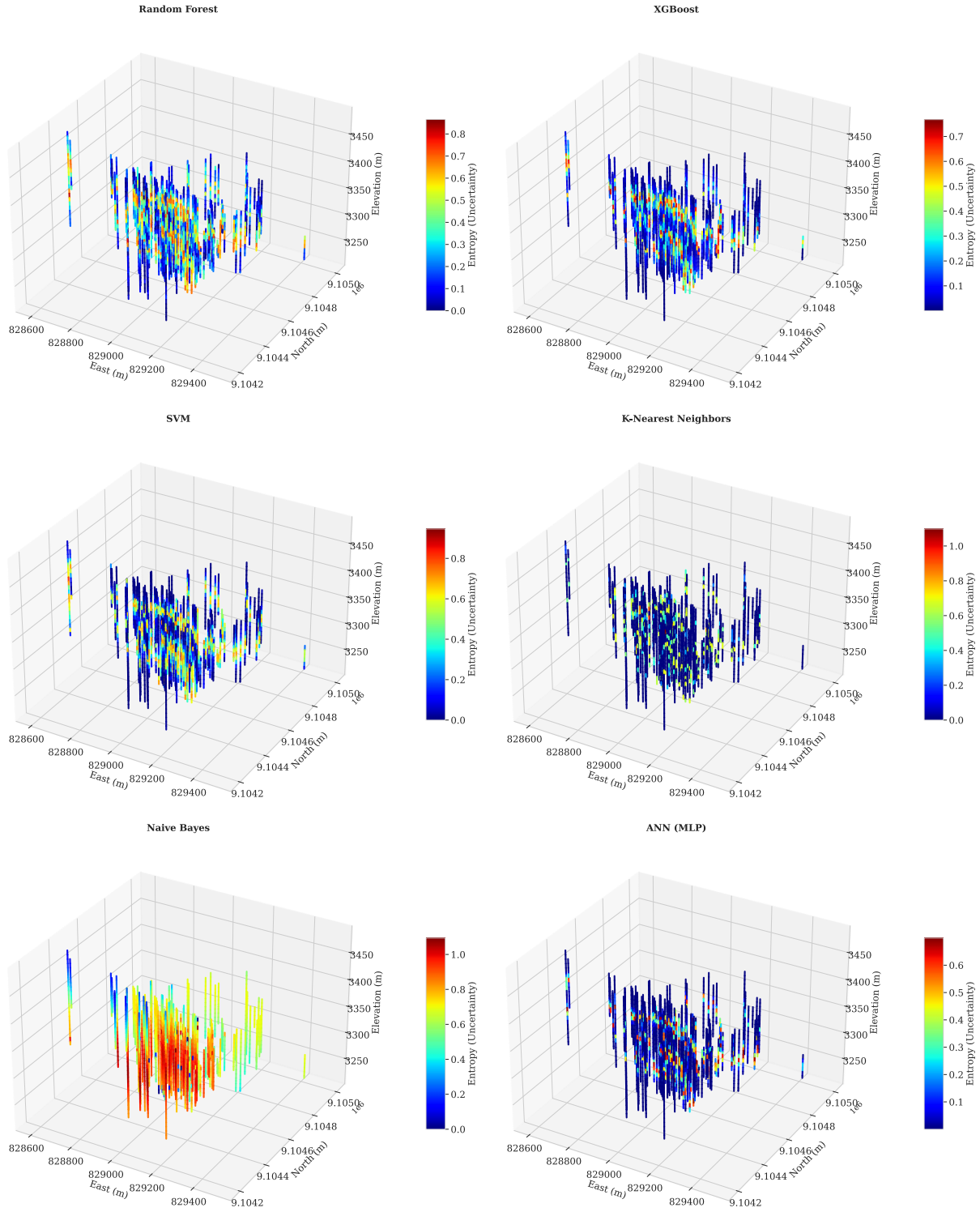

Figure 10. Three-dimensional visualization of predictive uncertainty along drill holes for each supervised classification model

From a geological standpoint, high-entropy areas may correspond to: (i) lithological transitions that are gradational or fault-disrupted, (ii) structural corridors with intense deformation, (iii) contact aureoles of intrusions with alteration halos, or (iv) zones with poor core recovery or sparse sampling density. These are precisely the domains that pose greater challenges for geological modeling and mineral prediction.

Thus, beyond its statistical interpretation, entropy becomes a powerful spatial indicator of geological uncertainty. Integrating this information into exploration workflows can enhance decision-making by flagging zones that merit further

geological mapping, sampling, or validation drilling. This approach aligns with previous recommendations by Giraud et al. [34], Fu et al. [16], and Ibrahim et al. [17], who emphasize the role of uncertainty analysis in identifying geologically complex and potentially mineralized targets that may otherwise be underestimated by deterministic models.

The performance distributions of each model under cross-validation are summarized in Figure 11. The ANN-MLP model achieved the highest average performance, closely followed by XGBoost, SVM, and Random Forest, whose accuracy ranges show considerable overlap. In contrast, Naive Bayes ranks significantly lower. This pattern was statistically confirmed using the non-parametric Friedman test ($X^2 = 20.98, p = 0.00082$), complemented by the Nemenyi post-hoc test, which revealed significant differences between Naive Bayes and the top-performing models ($p < 0.05$). These findings are consistent with Alzubaidi et al. [15], who recommend the use of neural network and boosting-based models for geological tasks due to their superior generalization capacity and ability to handle complex nonlinear relationships. These results indicate that ANN-MLP and XGBoost outperform other classifiers, especially in complex geological environments where nonlinear boundaries and high-dimensional feature spaces predominate. Their robustness and predictive accuracy make them excellent candidates for implementation in lithological modeling tasks, particularly in structurally heterogeneous deposits.

From a geological perspective, zones of high uncertainty (high entropy) tend to be concentrated along contacts between lithological units and in areas characterized by potential structural complexity or limited data density. These areas correspond to diffuse transition zones, where lithological boundaries may not be sharply defined due to gradational changes, the presence of structural features such as faults or shear zones, or sparse drill-core sampling. Higher entropy values are observed in the northeastern portion of the model block, which may be related to lithological variability associated with unmapped minor intrusions or complex diagenetic contacts. This type of analysis enables the prioritization of future sampling or drilling campaigns aimed at reducing uncertainty in key zones critical for geological characterization and resource evaluation.

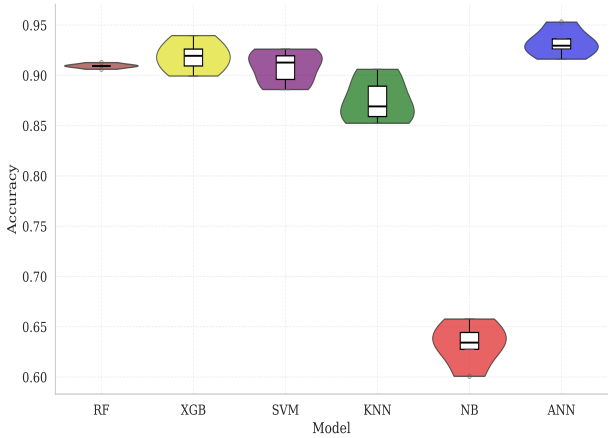

Figure 11. Comparison of machine learning model performance based on accuracy under cross-validation

4. Conclusions

This study demonstrated the strong potential of supervised machine learning models, particularly the ANN-MLP and XGBoost, for lithological classification based on spatial and geochemical drillhole data in a structurally complex gold deposit. These models achieved high predictive accuracy (F1-score > 94%) and showed greater robustness in classifying samples across lithological boundaries, outperforming traditional classifiers such as Naive Bayes, which failed to generalize in geologically ambiguous contexts. The integration of spatial coordinates significantly enhanced model coherence across the domain, confirming the importance of including positional variables in mineral exploration workflows. Moreover, the application of stratified cross-validation and spatially-aware data partitioning contributed to a more realistic and unbiased evaluation of model performance.

A major contribution of this research lies in the integration of entropy-based uncertainty analysis. By quantifying predictive confidence at the sample level, entropy served not only as a measure of model reliability but also as a tool for geological interpretation. High-entropy zones were consistently aligned with lithological transitions, structural discontinuities, and areas of intrusive contacts, which are typically characterized by complex mineralizing processes. The ability to visualize and quantify uncertainty provides practical advantages for exploration, including the identification of domains that require infill drilling, additional sampling, or re-logging. In this way, uncertainty-aware models offer a more transparent and risk-sensitive approach to subsurface

classification, moving beyond conventional accuracy-based assessments.

Future research should aim to extend this framework by incorporating additional geological variables such as structural features, alteration indices, or core imagery to enrich the feature space and reduce uncertainty. Likewise, the use of advanced architectures like graph neural networks (GNNs) is encouraged, as they can better capture spatial dependencies between samples and geological continuity along drillholes. Expanding the methodology to deposits with greater lithological heterogeneity and deformation complexity will also be essential to validate its robustness in diverse geological settings. In addition, future studies could explore class-level performance metrics (confusion matrices or per-class F1-scores) to gain more detailed insights into model behavior across different lithological units. Ultimately, combining machine learning with spatial uncertainty quantification opens new pathways for informed and geologically grounded decision-making in mineral exploration. This integrated approach not only improves the reliability of lithological models but also supports more effective drill targeting, optimizes resource allocation, and reduces geological risk, ultimately enhancing decision-making in exploration and development planning.

References

[1]. Li, H., Wan, B., Chu, D., Wang, R., Ma, G., Fu, J., & Xiao, Z. (2023). Progressive Geological Modeling and Uncertainty Analysis Using Machine Learning. *ISPRS International Journal of Geo-Information, 12*(3).

[2]. Akyildiz, O., Basarir, H., & Ellefmo, S. L. (2025). The development of a lithology prediction model using measurement while drilling data in a quartzite quarry. *International Journal of Mining, Reclamation and Environment, 39*(2), 93–109.

[3]. Saporetti, C. M., da Fonseca, L. G., & Pereira, E. (2019). A Lithology Identification Approach Based on Machine Learning With Evolutionary Parameter Tuning. *IEEE Geoscience and Remote Sensing Letters, 16*(12), 1819–1823.

[4]. Salehi, S. M., & Honarvar, B. (2014). Automatic Identification of Formation Lithology from Well Log Data: A Machine Learning Approach. *Journal of Petroleum Science Research, 3*(2), 73–82.

[5]. Min, X., Qiu, P., & Zhang, F. (2020). Research and application of logging lithology identification for igneous reservoirs based on deep learning. *Journal of Applied Geophysics, 173*.

[6]. Hatherly, P., Leung, R., Scheding, S., & Robinson, D. (2015). Drill monitoring results reveal geological conditions in blasthole drilling. *International Journal of Rock Mechanics and Mining Sciences, 78*, 144–154.

[7]. Kor, K., Ertekin, S., Yamanlar, S., & Altun, G. (2021). Penetration rate prediction in heterogeneous formations: A geomechanical approach through machine learning. *Journal of Petroleum Science and Engineering, 207*.

[8]. Shebl, A., Kusky, T., & Csámer, Á. (2022). Advanced land imager superiority in lithological classification utilizing machine learning algorithms. *Arabian Journal of Geosciences, 15*.

[9]. Akyildiz, O., Basarir, H., Vezhapparambu, V. S., & Ellefmo, S. (2023). MWD Data-Based Marble Quality Class Prediction Models Using ML Algorithms. *Mathematical Geosciences, 55*, 1059–1074.

[10]. Lee, H., & Lee, H. P. (2023). Formation lithology predictions based on measurement while drilling (MWD) using gradient boosting algorithms. *Geoenergy Science and Engineering, 227*.

[11]. Casado-Vara, R., Martín-Luis, M. C., Casillas-Ruiz, R., Valero-Garcés, B. L., & Jambrina-Enríquez, M. (2025). A comparative machine learning study for lithology classification in volcano-fluvial deposits from La Vega de La Laguna (Tenerife, Canary Islands): Application to core lithology log using all-elements and major elements approach. *Applied Computing and Geosciences, 27*, 100273.

[12]. Komadja, G. C., Westman, E., Rana, A., & Vitalis, A. (2025). A Machine Learning Approach to Lithology Classification in Mining Using Measurement While Drilling and Exploration Data. *Mining, Metallurgy & Exploration, 42*(4), 1955–1973.

[13]. Samir, K., El-Sharkawi, M., El-Barkooky, A. N., & Hammed, M. S. (2023). Application of Remote Sensing in Lithological Mapping of Umm Tawat Precambrian Rock Assemblage, North Eastern Desert, Egypt. *Journal of Mining and Environment, 14*(2), 473–491.

[14]. Xiao, K., Jiao, C., Zou, C., Yang, Y., Li, H., Hu, X., Chen, M., Yin, D., Yang, P., & He, R. (2026). Evaluation of machine learning methods for lithology classification of sandstone-type uranium deposit based on well logging data in the Songliao Basin, Northeast China. *Nuclear Engineering and Technology, 58*(2), 103946.

[15]. Alzubaidi, F., Mostaghimi, P., Swietojanski, P., Clark, S. R., & Armstrong, R. T. (2021). Automated lithology classification from drill core images using convolutional neural networks. *Journal of Petroleum Science and Engineering, 197*, 107933.

[16]. Fu, D., Su, C., Wang, W., & Yuan, R. (2022). Deep learning based lithology classification of drill core images. *PLOS ONE, 17*(7), e0270826.

[17]. Ibrahim, A. F., Ahmed, A., & Elkatatny, S. (2023). Applications of Different Classification Machine

Learning Techniques to Predict Formation Tops and Lithology While Drilling. *ACS Omega, 8*(45), 42152–42163.

[18]. Li, Q., Chi, P., Fu, J., Zhang, X., Yu, S., Zhang, C., Wang, P., Feng, C., & Pan, Y. (2023). A comprehensive machine learning model for lithology identification while drilling. *Geoenergy Science and Engineering, 231*, 212333.

[19]. Abdullah, M. A. M., Mohammed, A. A., & Awad, S. R. (2024). RockDNet: Deep Learning Approach for Lithology Classification. *Applied Sciences, 14*(13), 5511.

[20]. Rohit, Manda, S. R., Raj, A., Dheeraj, A., Rawat, G. S., & Choudhury, T. (2024). Identification of Lithology from Well Log Data Using Machine Learning. *EAI Endorsed Transactions on Internet of Things, 10*, 5634.

[21]. Zhang, P., Gao, T., Li, R., & Fu, J. (2024). Advanced Machine Learning Framework for Enhanced Lithology Classification and Identification. *In International Petroleum Technology Conference 2024* (pp. 4318–4330).

[22]. Bressan, T. S., Kehl de Souza, M., Girelli, T. J., & Junior, F. C. (2020). Evaluation of machine learning methods for lithology classification using geophysical data. *Computers & Geosciences, 139*, 104475.

[23]. Burak, T., Sharma, A., Hoel, E., Kristiansen, T. G., Welmer, M., & Nygaard, R. (2024). Real-Time Lithology Prediction at the Bit Using Machine Learning. *Geosciences, 14*(10), 250.

[24]. Marquina-Araujo, J. J., Cotrina-Teatino, M. A., Cruz-Galvez, J. A., Noriega-Vidal, E. M., & Vega-Gonzalez, J. A. (2024). Application of Autoencoders Neural Network and K-Means Clustering for the Definition of Geostatistical Estimation Domains. *Mathematical Modelling of Engineering Problems, 11*(5), 1207–1218.

[25]. Li, K., Ren, T., Yao, N., Roberts, J., Song, H., Li, Z., & Liang, C. (2025). Real-time lithology identification while drilling based on drilling parameters analysis with machine learning. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources, 11*, 44.

[26]. Peng, Q., Wang, Z., Wang, G., Zhang, W., Chen, Z., & Liu, X. (2023). 3D Mineral Prospectivity Mapping from 3D Geological Models Using Return–Risk Analysis and Machine Learning on Imbalance Data. *Minerals, 13*(11), 1384.

[27]. Xiang, J., Xiao, K., Carranza, E. J. M., Chen, J., & Li, S. (2020). 3D Mineral Prospectivity Mapping with Random Forests: A Case Study of Tongling, Anhui, China. *Natural Resources Research, 29*(1), 395–414.

[28]. Khalifa, H., Tomomewo, O. S., Ndulue, U. F., & Berrehal, B. E. (2023). Machine Learning-Based Real-Time Prediction of Formation Lithology and Tops Using Drilling Parameters with a Web App Integration. *Eng, 4*(3), 139.

[29]. Maleki, M., Mery, N., Soltani-Mohammadi, S., Plaza-Carvajal, J., & Varouchakis, E. A. (2025). Integrating Geological Domains into Machine Learning for Ore Grade Prediction: A Case Study from a Porphyry Copper Deposit. *Minerals, 15*(11), 1175.

[30]. Günther, C., Simán, F., Mokayed, H., Liwicki, M., Jansson, N., McDonnell, P., Hermansson, T., & Liwicki, F. S. (2025). Machine learning for drill core image analysis: A review. *Ore Geology Reviews, 187*, 106974.

[31]. Jordão, H., Sousa, A. J., & Soares, A. (2023). Using Bayesian Neural Networks for Uncertainty Assessment of Ore Type Boundaries in Complex Geological Models. *Natural Resources Research, 32*(6), 2495–2514.

[32]. Nie, X., Lu, C., & Luo, K. (2023). Uncertainty assessment of 3D geological models based on spatial diffusion and merging model. *Open Geosciences, 15*(1), 20220456.

[33]. Pirot, G., Joshi, R., Giraud, J., Lindsay, M. D., & Jessell, M. W. (2022). loopUI-0.1: indicators to support needs and practices in 3D geological modelling uncertainty quantification. *Geoscientific Model Development, 15*, 4689–4717.

[34]. Giraud, J., Lindsay, M., Jessell, M., & Ogarko, V. (2020). Towards plausible lithological classification from geophysical inversion: Honouring geological principles in subsurface imaging. *Solid Earth, 11*(2), 419–436.

[35]. Colquehuanca, E. (2016). Planeamiento de minado a largo plazo haciendo uso del software Whittle 4.4 Cía. Minera Aurífera Santa Rosa Comarsa S.A. [Tesis de ingeniería]. Universidad Nacional de San Agustín de Arequipa.

[36]. Montoya, D. (2010). Geología del depósito aurífero Santa Rosa (CNM7-010, 8 pp.). INGEMMET.

[37]. Montoya, D. E., Noble, D. C., Eyzaguirre, V. R., & Desrosiers, D. F. (1995). Sandstone-hosted gold deposits: a new exploration target is recognized in Peru. *Engineering & Mining Journal, 196*(6), 34–41.

[38]. Cotrina-Teatino, M. A., Marquina-Araujo, J. J., & Riquelme, Á. I. (2025). Comparison of Machine Learning Techniques for Mineral Resource Categorization in a Copper Deposit in Peru. *Natural Resources Research, 34*(4), 1987–2005.

[39]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

[40]. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

[41]. Trott, M., Leybourne, M., Hall, L., & Layton-Matthews, D. (2022). Random forest rock type classification with integration of geochemical and photographic data. *Applied Computing and Geosciences, 15*, 100090.

[42]. Silversides, K. L., & Melkumyan, A. (2022). Machine learning for classification of stratified geology from MWD data. *Ore Geology Reviews, 142*, 104737.

[43]. Lin, X., & Yin, S. (2023). Lithology identification based on interpretability integration learning. *Earth Science Informatics, 16*(3), 2211–2222.

[44]. Madani, A. A. (2024). Evaluation of Band Ratio Technique for Prediction of Iron-Titanium Mineralization Using Ensemble Machine Learning Model: A Case Study from Khamal area, Western Saudi Arabia. *Journal of Mining and Environment, 15*(4), 1357–1371.

[45]. Gao, M., Wang, G., Xu, Y., Mou, N., Huang, L., Zuo, L., & Wu, R. (2023). 3D mineral exploration Cu-Zn targeting with multi-source geoscience datasets in the Weilasituo-bairendaba district, Inner Mongolia, China. *Frontiers in Earth Science, 11*, 1102640.

[46]. Wei, Y., Xing, Z., Jian, C., Wang, K., Wu, S., & Chiam, K. (2021). Use of Tree-Based Machine Learning Methods for Stratigraphic Classification in 3D Geological Modelling. *IOP Conference Series: Earth and Environmental Science, 861*(7), 072039.

[47]. Hu, Y., Wang, Z. Z., Guo, X., Kek, H. Y., Ku, T., Goh, S. H., Leung, C. F., Tan, E., & Zhang, Y. (2024). Three-dimensional reconstruction of subsurface stratigraphy using machine learning with neighborhood aggregation. *Engineering Geology, 337*, 107588.

[48]. Dong, S.-Q., Zhong, Z.-H., Cui, X.-H., Zeng, L.-B., Yang, X., Liu, J.-J., Sun, Y.-M., & Hao, J.-R. (2023). A deep kernel method for lithofacies identification using conventional well logs. *Petroleum Science, 20*, 1411–1428.

[49]. Zhang, J., He, Y., Zhang, Y., Li, W., & Zhang, J. (2022). Well-Logging-Based Lithology Classification Using Machine Learning Methods for High-Quality Reservoir Identification: A Case Study of Baikouquan Formation in Mahu Area of Junggar Basin, NW China. *Energies, 15*(10), 3675.

[50]. Lv, P., Chen, W., & Zou, Y. (2025). Geological Target Recognition Method Based on Multi-mode Adaptive Prediction System: Algorithms and Applications. *Natural Resources Research, 34*(6), 2973–2992.

[51]. Zuo, R., & Carranza, E. J. M. (2023). Machine Learning-Based Mapping for Mineral Exploration. *Mathematical Geosciences, 55*(7), 891–895.

[52]. Cotrina-Teatino, M., Marquina-Araujo, J., Noriega-Vidal, E., Mamani-Quispe, J., Ccatamayo-Barrios, J., Gonzalez-Vasquez, J., & Arango-Retamozo, S. (2024). Predicting open pit mine production using machine learning techniques: A case study in Peru. *Journal of Mining and Environment, 15*(4), 1343–1355.

[53]. Horrocks, T., Holden, E. J., & Wedge, D. (2015). Evaluation of automated lithology classification architectures using highly-sampled wireline logs for coal exploration. *Computers & Geosciences, 83*, 86–96.

[54]. Xie, Y., Zhu, C., Zhou, W., Li, Z., Liu, X., & Tu, M. (2018). Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. *Journal of Petroleum Science and Engineering, 160*, 182–193.

[55]. Chen, G., Chen, M., Hong, G., Lu, Y., Zhou, B., & Gao, Y. (2020). A new method of lithology classification based on convolutional neural network algorithm by utilizing drilling string vibration data. *Energies, 13*(4), 888.

[56]. Shi, Y., Liao, J., Gan, L., & Tang, R. (2024). Lithofacies Prediction from Well Log Data Based on Deep Learning: A Case Study from Southern Sichuan, China. *Applied Sciences, 14*(18), 8195.

[57]. Soulaimani, S., Soulaimani, A., Abdelrahman, K., Miftah, A., Fnais, M. S., & Mondal, B. K. (2024). Advanced machine learning artificial neural network classifier for lithology identification using Bayesian optimization. *Frontiers in Earth Science, 12*, 1473325.

[58]. Mousavi, S. H. R., & Hosseini-Nasab, S. M. (2024). A novel approach to classify lithology of reservoir formations using GrowNet and Deep-Insight with physic-based feature augmentation. *Energy Science & Engineering, 12*(10), 4453–4477.

[59]. Cotrina, M., Marquina, J., Mamani, J., Arango, S., Gonzalez, J., Ccatamayo, J., & Noriega, E. (2024). Predictive model using machine learning to determine fuel consumption in CAT-777F mining equipment. *International Journal of Mining and Mineral Engineering, 15*(2), 147–160.

[60]. Stock, A. (2025). Choosing blocks for spatial cross-validation: lessons from a marine remote sensing case study. *Frontiers in Remote Sensing, 6*, 1531097.

[61]. Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling, 411*, 108815.

[62]. Widodo, S., Brawijaya, H., & Samudi, S. (2022). Stratified K-fold cross validation optimization on machine learning for prediction. *Sinkron, 6*(4), 2407–2414.

[63]. Adeli, A., Emery, X., & Dowd, P. (2018). Geological modelling and validation of geological interpretations via simulation and classification of quantitative covariates. *Minerals, 8*(1), 7.

[64]. Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. *Annals of Mathematics and Artificial Intelligence, 81*(3), 429–450.

[65]. Markoulidakis, I., & Markoulidakis, G. (2024). Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis. *Technologies, 12*(7), 113.

[66]. Mishra, S., & Ayyub, B. M. (2019). Shannon Entropy for Quantifying Uncertainty and Risk in Economic Disparity. *Risk Analysis, 39*(10), 2160–2181.

[67]. Alam Mohd., Siddiqui, A. N., Shamim, S. K., Ahmad, A., Faiz, M. (2025). A comparative evaluation of geostatistical techniques of frequency ratio (FR), Shannon Entropy (SE), and analytical hierarchy (AHP) in landslide prediction: a case study of Uttarakhand, India. *Discover Geoscience, 3*, 28.

[68]. Robinson, D. W. (2008). Entropy and Uncertainty. *Entropy, 10*(4), 493–506.

[69]. Polo Salinas, J., Marquina Araujo, J. J., & Cotrina Teatino, M. A. (2025). Uncertainty in underground mining operations: a bibliometric and systematic literature review analysis. *World Journal of Engineering, 22*(6), 1328–1343.

[70]. Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *In Advances in Neural Information Processing Systems 25 (NeurIPS 2012)* (pp. 2951–2959).

[71]. Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*, 281–305.

[72]. Lin, N., Fu, J., Jiang, R., Li, G., & Yang, Q. (2023). Lithological Classification by Hyperspectral Images Based on a Two-Layer XGBoost Model, Combined with a Greedy Algorithm. *Remote Sensing, 15*(15), 3764.

[73]. Shashel, A. (2022). Uncertainty analysis of Supervised machine learning predictions applied to Lithology classification. Master's thesis, University of Stavanger.