POLITECNICO DI TORINO
Repository ISTITUZIONALE

Evaluating StackingC and ensemble models for enhanced lithological classification in geological mapping

Original
Evaluating StackingC and ensemble models for enhanced lithological classification in geological mapping / Farhadi, Sasan; Tatullo, Samuele; Boveiri Konari, Mina; Afzal, Peyman. - In: JOURNAL OF GEOCHEMICAL EXPLORATION. - ISSN 0375-6742. - 260:(2024), pp. 1-14. [10.1016/j.gexplo.2024.107441]

Availability:
This version is available at: 11583/2986849 since: 2024-03-12T08:25:12Z

Publisher:
Elsevier

Published
DOI:10.1016/j.gexplo.2024.107441

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

16 May 2024

Journal of Geochemical Exploration 260 (2024) 107441

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/gexplo

Evaluating StackingC and ensemble models for enhanced lithological classification in geological mapping

Sasan Farhadi [a,*], Samuele Tatullo [b], Mina Boveiri Konari [c], Peyman Afzal [d]

[a] Department of Structural, Geotechnical and Building Engineering, Polytechnic University of Turin, Turin, Italy
[b] IMT School for Advanced Studies, Lucca, Italy
[c] Samaneh Kansar Zamin Company, Tehran, Iran
[d] Department of Petroleum and Mining Engineering, South Tehran Branch, Azad University, Tehran, Iran

ARTICLE INFO

Keywords:
Machine learning
Ensemble learning
Boosting
Stacking
Bagging
Lithological classification

ABSTRACT

Lithological classification is a crucial aspect of mineral exploration, providing insights into rock and mineral types in a given area. Conventional methods for lithological classification can be limited in terms of coverage, accuracy, and efficiency, often experiencing significant time and cost. Machine learning techniques have demonstrated considerable potential in improving the efficiency and accuracy of this process. In this study, the effectiveness of ensemble learning models, including boosting, stacking, and bagging, was compared to logistic regression (LR) and support vector machines (SVM) as baseline models for predicting lithological classes using geochemical and geological data. Notably, the stackingC model, a novel stacking variant, stood out as the best-performing model. It achieved remarkable Cohen's Kappa and Matthews Correlation Coefficient (MCC) scores of 97.10% and 93.70%, respectively. The Bagged Decision Trees and Adaboost models also demonstrated strong performance, with a kappa score of 97.10% and an MCC of 92.80%. In contrast, the LR model underperformed, scoring 37.70% in kappa and 43% in MCC. These results emphasize the potential of ensemble learning models for lithological classification, mainly when dealing with complex, nonlinear relationships between input variables and output labels. Such models hold promise for improving accuracy and generalization in mineral exploration.

1. Introduction

Lithological classification has a key role in mineral exploration, as it provides valuable information about the type, quality, and distribution of rocks and minerals in a given area. Accurate lithological classification can help identify potential mineral deposits and optimize exploration and mining activities. However, traditional methods for lithological classification, such as geological mapping and drilling, are often limited in their coverage and accuracy and can be expensive and time-consuming. Therefore, there is a growing need to identify new methods for lithological classification that can improve the accuracy and efficiency of mineral exploration. By forecasting rock unit lithology, it becomes possible to identify not only the name and distribution of rock units in a limited area but also examine different types of alterations and the association of ore minerals. This could significantly enhance our understanding of geological formations and provide valuable insights for mineral exploration and mining activities.

In recent years, the application of machine learning (ML) techniques has gained attention in various fields, in particular, geoscience. Many researchers in this field have been implementing ML techniques to improve the accuracy and efficiency of lithological classification Dev and Eden (2019). Harris and Grunsky (2015) employed the Random Forest (RF) method for lithology mapping in Canada, utilizing lake sediment geochemical, airborne gamma-ray spectrometer, and magnetic data. Their study reveals that RF classification can generate reliable predictive lithologic maps, making it a beneficial tool for complementing field mapping activities in poorly mapped regions.

Bressan et al. (2020) used four machine learning techniques (Multilayer Perceptron, decision tree, RF, and support vector machine (SVM) to analyze three standard data templates and a practical data template in a lithological classification problem for multivariate log parameter data from offshore wells from the International Ocean Discovery Program (IODP). The authors found that random forest characteristics allowed for better results in lithological classification, leading to accurate results with small statistical variances. Gifford and Agah (2010) employed multi-agent machine learning and classifier combination to

* Corresponding author.
E-mail address: sasan.farhadi@polito.it (S. Farhadi).

https://doi.org/10.1016/j.gexplo.2024.107441
Received 19 June 2023; Received in revised form 4 February 2024; Accepted 23 February 2024
Available online 6 March 2024
0375-6742/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

classify rock facies sequences using wireline well log data. The paper focuses on constructing successful sets of classifiers that collaborate to increase classification accuracy. Akkas et al. (2015) showed that decision tree algorithms stand as an accurate and rapid method for mineral classification/identification using characteristic X-ray intensities produced in a typical SEM-EDS, achieving high accuracy rates for training sets with 800 data, even though by using non-standard quality training sets.

Lima et al. (2020) employed deep convolutional neural networks and pre-trained models to expedite and enhance microfacies classification in petrographic thin sections. The authors utilized parallel polarized petrographic thin-section images with 10× magnification to differentiate between classes based on grain size. The authors found that their approach achieves high accuracy for microfacies classification when using images from the Sycamore Formation and comparable accuracy to a petrographer. SVM algorithms were also used by Yu et al. (2012) to automate the lithological classification of an area in northwestern India, building a first-pass lithological map with high accuracy, and by Abedi et al. (2012) to select the best area for exploratory drilling, building a prospectivity map for the Now Chun copper deposit located in Iran, reducing the amount of risk for managers of exploratory projects in continuing their task. Choubin et al. (2023) employed the SVM model to predict land susceptibility to dust emission while simultaneously assessing the significance of critical factors driving dust emissions within the study area.

In this paper, ensemble learning techniques such as boosting, stacking, and bagging models, along with some of their modifications, were utilized to classify lithological formations using geochemical and geospatial data. The ensemble methods were chosen for their capability to reduce the variance and improve the accuracy of machine learning techniques for classification. The performance of these models was compared to commonly used models, logistic regression (LR) (Felicisimo et al., 2013; Das et al., 2010; Harris and Pan, 1999) and SVM, and provide insights into the interpretability of the proposed models. This study contributes to the growing body of research on using machine learning techniques for geoscience applications and highlights the potential of ensemble learning methods for lithological classification.

2. Geological setting

The study area is located southeast of Esfahan province, between the Sanandaj-Sirjan and Central Iran structural zones (Fig. 1). This region is characterized by diverse ore deposit types, including epithermal, porphyry, massive sulfide, manto, and others associated with the Urumieh-Dokhtar and Naein ophiolitic complexes. These deposits were formed due to the evolution of the Neotethys Ocean from opening to subduction and, finally, an orogenic event (Hassanzadeh and Wernicke, 2022; Dercourt et al., 1986). The ore deposits are hosted within Cretaceous volcano-sedimentary sequences that encompass several deposits such as the Kahang porphyry copper deposit, Qalehdar, Zafarghand, Kalchuyeh epithermal Cu, Feyzabad sediment-hosted Zn—Pb deposit, volcano-sedimentary Mn (Cu) Benvid, Kachumesghal volcanic copper and Meskat deposit (Heidari et al., 2023). The distribution of andesite and basaltic andesite in the area is notable, with copper mineralization also occurring in these units (Movahednia et al., 2022).

According to the field observation and collected samples, a geological map of the study area was prepared at a 1:5000 scale. The study area primarily consists of rock units from the Cretaceous period, which are a combination of volcanic and sedimentary rocks (Fig. 2). The volcanic rocks primarily consist of basaltic lava, andesite, basaltic andesite, rhyolite, and rhyolitic tuff. The sedimentary rocks are mainly Cretaceous limestone and dolostone. Cenozoic units are less widespread in the region. Diorite to granodiorite rocks intruded into the Cretaceous rocks. The subsequent section provides a concise discussion of the observed characteristics of the rock units within the studied area (Fig. 3).

k^{bt} unit: this rock unit has limited outcrops in the northwestern part of the area. Its primary lithology consists of olive green to dark basalts, and to some extent andesite to andesitic basalt. The fractures in these rocks are often filled with iron oxide. In some locations, small outcrops of green-colored pyroclastic rocks, including tuff and tuffite, can be

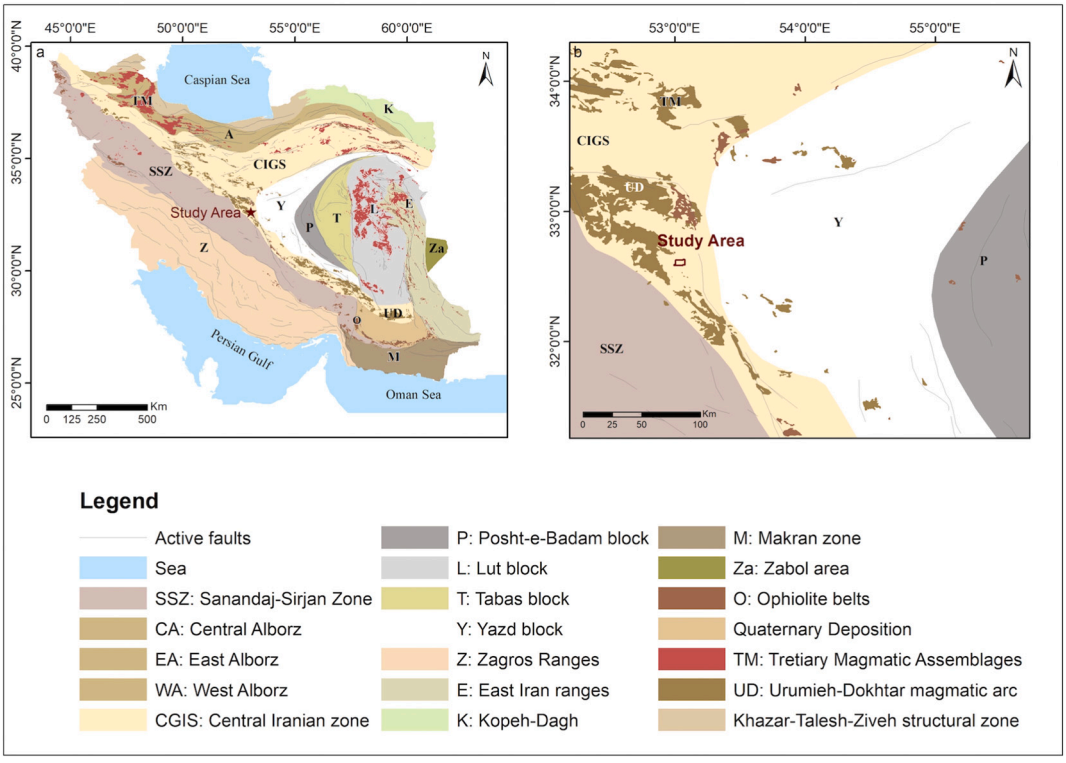

Fig. 1. Left: Study area located within the structural zone of Iran. Right: Close-up view of the study area in Central Iran.

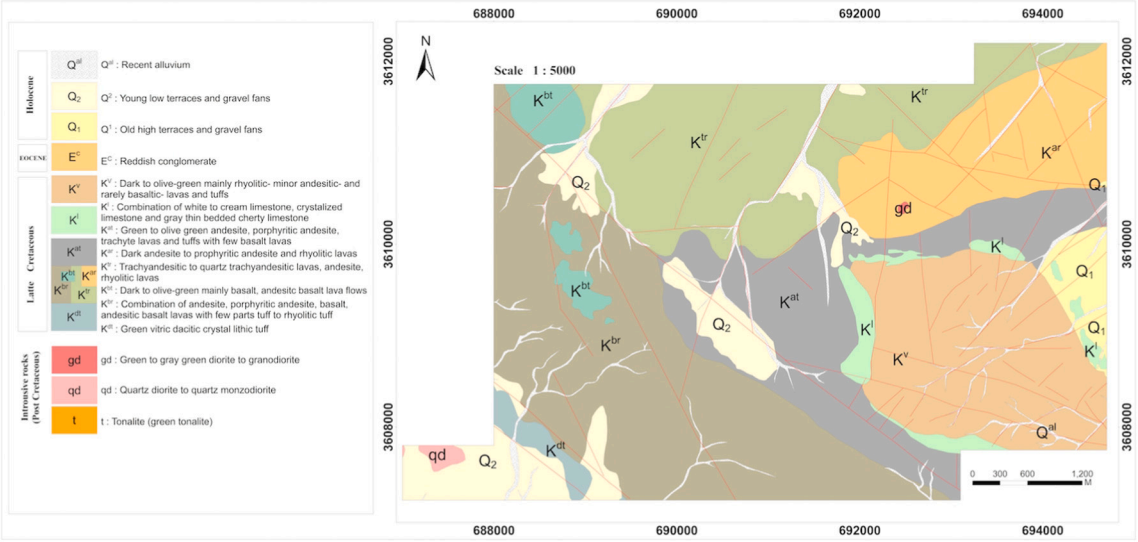

Fig. 2. Geological map of the study area at a scale of 1:5000.
(Source: Samaneh Kansar Zamin, 2023).

observed. This unit is also exposed within the k^{br} unit in patches.

k^{tr} unit: another unit in the area is the K^{tr}, which is attributed to the Cretaceous period. It has its most extensive outcrop in the northern half of the study area. The dominant lithology of this unit includes trachyte-andesite lavas, quartz trachyte-andesite, andesite, and rhyolitic lavas. Basaltic lavas and pyroclastic deposits are less common. The unit has a green outcrop due to the presence of chlorite. Its upper boundary corresponds to Quaternary sediments and the K^{l} and K^{at} units.

k^{ar} unit: this rock unit is exposed in the northeastern part of the area with an east-west trend. It is primarily composed of dark green to olive-colored andesite to rhyolite lavas. Intervals of tuff sediments can be observed among the volcanic units, which have undergone minor alteration. Diorite to granodiorite intrusive rocks have penetrated parts of this unit, resulting in traces of propylitic alteration at their contact. This unit has two types of siliceous veins: mineralized and non-mineralized. In some locations, the joints and fractures of the veins are filled with iron oxide and pyrite molds. These veins have a northwest-southeast trend and vary in length from 10 to 100 m and in thickness from 20 to 30 cm.

k^{at} unit: this rock unit has a general northeast-southwest trend and is most widespread in the eastern and central parts of the exploration area. It is composed of andesite, andesite porphyry, trachytic lavas, tuff, and to a lesser extent, basaltic lavas. The outcrops are mainly green to olive green in color. This rock unit has undergone prophylactic and argillic alterations. Quaternary terraces are exposed at its southern contact. Thinly layered gray-green tuff can be observed as interlayers within the andesite units. It is worth noting that copper mineralization has occurred in some parts of this unit in the form of malachite and chalcopyrite as open-space filling.

k^{l} unit: this lithology comprises biomicrite limestones, crystallized limestones, and chert limestones. It is most widely distributed in the eastern and southeastern parts of the exploration area. The outcrops are visible in white to light cream color. Cretaceous volcanic-pyroclastic units limit the northern and southern contacts of the k^{l} unit. Based on field observations, limestone layers are exposed as thin, medium, and thick. In some locations, the limestones are dolomitic and heavily crushed, while in others, they are foliated and partially oriented. Scattered copper mineralization is present as malachite and chalcopyrite within and sometimes at the contact of limestone units with andesite rocks. Old carvings and goethite zones can also be observed in this unit.

k^{v} unit: this rock unit is the primary host of ore mineralization and has a generally northwest-southeast direction, like other volcanic units. The dominant lithology consists of olive green to dark rhyolitic lavas, which cover almost all of the range's eastern parts. Additionally, this unit contains a combination of andesite and basalt with interlayers of tuff. The basalt outcrops are darker in color and tend towards black. Mineralized outcrops are scattered within the K^{v} units, measuring about 1 to 5 m in length and several centimeters in thickness. These patches are small in size but numerous. Malachite mineralization has also been observed in this unit.

gd unit: the gd intrusive rocks, composed of diorite to granodiorite, are exposed in the northern part of the region. They have limited expansion and have intruded within older volcanic units, mainly of Cretaceous age, indicating that they are younger than these units. This mass has a mild morphology and is characterized by green and sometimes gray hills in the area.

qd unit: this unit is exposed northwest of the area and consists of quartz diorite to quartz monzonite with a gray-to-green color. It cuts through Lower Cretaceous volcanic rocks and has small outcrops in contact with volcanic and Quaternary deposits. The texture of this unit is granular.

Quaternary units: based on the geological map and field operations, the most extensive and youngest lithological unit in the area is the Quaternary sediments and terraces. These are exposed in all the range and include Q^{1}, consisting of alluvial alluviums and old river terraces exposed in the eastern parts; Q^{2}, including alluviums and new river terraces; and Q^{al}, including river sediments and recent alluviums.

3. Material and methods

3.1. Dataset

To prepare the geological map at a scale of 1:5000, 280 samples were carefully selected to represent all rock types, alteration products, and ore mineralization in the study area (Section 2). Of these samples, 215 samples were subjected to ICP-OES and fire assay analyses, providing required information on elemental compositions and mineral content. Additionally, 33 samples comprising 15 thin sections (TS) and 18 thin polished sections (TP) were subjected to microscopic examination. For a more in-depth understanding of alteration characteristics, 18 samples were analyzed using X-ray diffraction (XRD), while 10 samples were examined using X-ray fluorescence (XRF) techniques. The dataset was collected from various field observations and analytical results, combining geological knowledge with laboratory analyses to form a

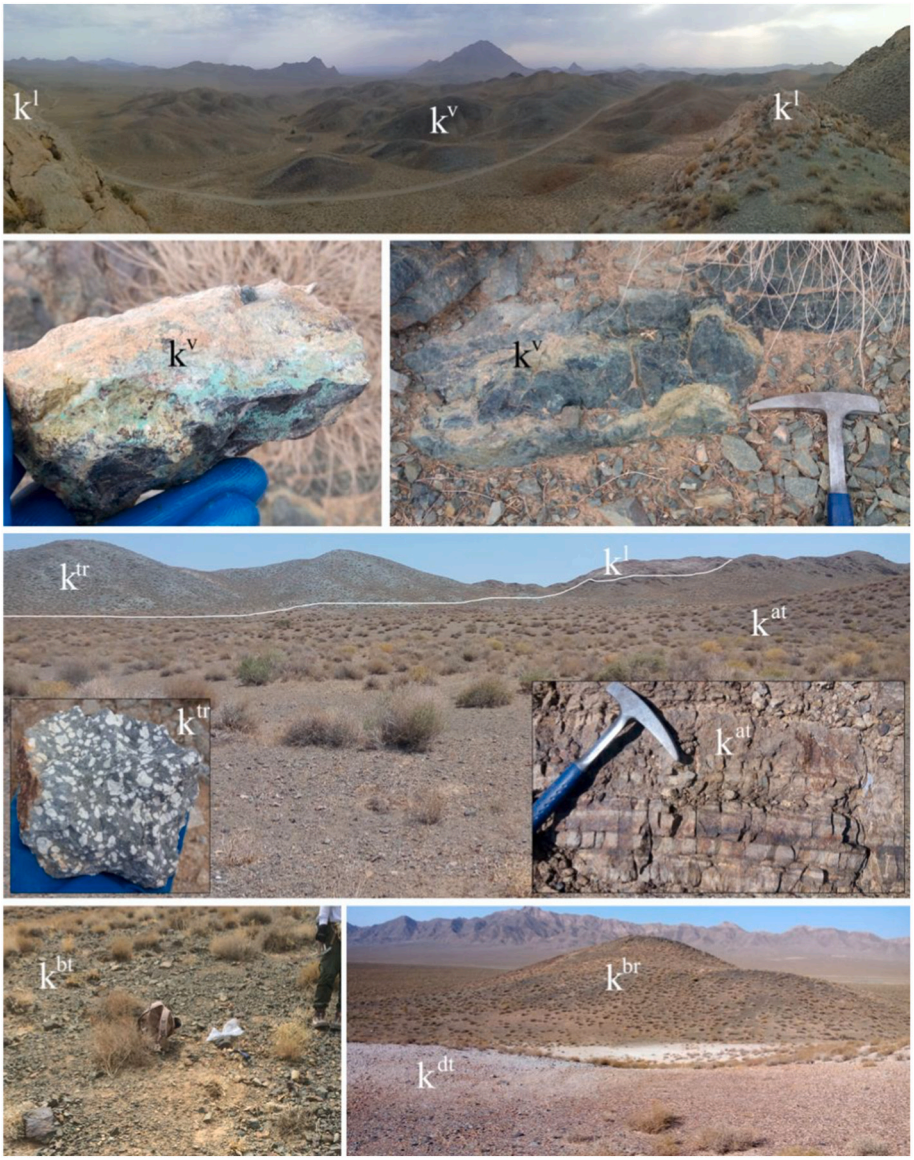

Fig. 3. Outcrop and hand specimen images from rock unit lithologies distributed in the study area.

comprehensive representation of the lithological diversity in the study area.

Table 1 presents the description of typical samples used in this study. Moreover, the predictive features utilized in model training include geological and geochemical parameters, including major (e.g., copper, lead, and zinc) and minor (e.g., manganese and sulfur) elements analyzed through ICP-OES and fire assay techniques, precious elements such as gold content, trace elements like molybdenum and antimony, and XRD patterns. These features have been thoughtfully selected to align with the research objectives and enhance the accuracy of lithological classification. The comprehensive list of features used for

Table 1
Description of the lithological samples.

Rock type	Fire assay	ICP-OES	XRF	TS	TP	XRD
k_{ar}	7	6	0	1	4	0
k_{at}	17	19	1	4	4	2
k_{v}	20	33	0	5	7	0
k_{tr}	38	29	8	1	3	16
k_{l}	13	13	1	1	0	0
k_{dt}	3	3	0	0	0	0
k_{bt}	1	0	0	1	0	0
k_{br}	0	0	0	1	0	0
g_{d}	0	0	0	1	0	0

Table 2
Utilized features for lithological classification.

Features	Description	Relevance to lithological characteristics
Major elements (ICP-OES)	Cu, Pb, Zn (base metal) content	Significant in ore mineralization
Major elements (ICP-OES)	Fe content	Indicate presence of iron-rich minerals
Precious elements (fire assay)	Au content	Significant in ore mineralization
Minor elements (ICP-OES)	Mn and S content	Indicate the source of ore-bearing fluid
Trace elements (ICP-OES)	Mo and Sb content	Significant in ore mineralization
XRD patterns	X-ray diffraction peak intensity	Indicate minerals and alteration product minerals
Microscopic studies	Texture and rock lithology investigation	Identification of rock lithology and ore minerals

prediction in this research is presented in Table 2, along with their relevance to lithological characteristics. Major elements obtained through ICP-OES, such as copper, lead, and zinc, are significant indicators of the base metal presence and are essential in identifying ore mineralization. Likewise, the iron content derived from ICP-OES analysis serves as an indicator of the presence of iron-rich minerals. Gold content elements determined through a fire assay play a key role in identifying valuable ore mineralization. Furthermore, minor elements such as manganese and sulfur content, also obtained via ICP-OES, provide insights into the source of ore-bearing fluids. Trace elements, including molybdenum and antimony content, are significant in ore mineralization. The XRD patterns, which measure X-ray diffraction peak intensity, are essential in identifying specific minerals and altering product minerals in geological formations. Additionally, microscopic studies containing texture and rock lithology reserve in identifying rock lithology and ore minerals further enrich the dataset's lithological insights. Fig. 4 illustrates the spatial distribution of the selected samples across the study area, providing a better understanding of their geological locations within the research area.

3.2. Machine learning

In 1959, Arthur Samuel coined the term *machine learning* and wrote the first known program under this category: the "Checkers Playing Program". Albeit inferential statistics was already a well-established research field since Arthur Samuel started the machine learning era, this subject started to conquer almost all branches of studies, helping researchers to understand better and uncover problems. Machine learning is a collection of methodologies that use computers' power to analyze billions of data in a short time. Machine learning is used in a wide variety of areas such as image segmentation (Seo et al., 2020; Xu et al., 2019; Arganda-Carreras et al., 2017), natural language processing (Wolf et al., 2020; Otter et al., 2021), sound event detection (Mesaros et al., 2021; Farhadi et al., 2024), identification of diseases in healthcare (Ngiam and Khor, 2019; Liu et al., 2023; Wang et al., 2023), computer vision (Stefenon et al., 2022), environmental science (Choubin and Rahmati, 2021; Mosavi et al., 2020) mineral exploration and anomaly detection (Farhadi et al., 2022; Maitre et al., 2019; Rodriguez-Galiano et al., 2015), and the list could continue given its nowadays spread usage.

Machine learning techniques can be divided according to the usage objectives; the principal reasons usually are making predictions or clustering the data or by the type of algorithm used; in this case, the classification is supervised, unsupervised, and reinforcement learning techniques. In a supervised algorithm, data are labeled, and the algorithm is capable of "seeing" these labels; this capacity helps computers and researchers to understand the performance of the algorithm. On the contrary, in an unsupervised algorithm, data do not present labels; this could be due to a lack of information or to a choice of the researcher; in this case, there have been created formulas that allow understanding the goodness of the algorithm. In reinforcement learning techniques, the computer should be considered a pure learner, meaning it learns through experience and failures. Examples of this algorithm are common nowadays, but probably the most famous one is AlphaGo.

In this comparative study, an extensive analysis is undertaken to evaluate ensemble techniques such as boosting, stacking, and bagging. Ensemble methods are one of the most robust techniques in the supervised ML approach that combine many simple "building block" models to obtain a single and hopefully more powerful model (Dietterich, 2000; Sagi and Rokach, 2018).

3.3. Boosting

Boosting, as introduced by Schapire (1990), is a powerful ensemble technique that creates a robust classification model using weak classifiers. A weak classifier is one that performs slightly better than a random choice in terms of error rate. Therefore, boosting is just the application of a sequence of weak classification algorithms, say $f_m(\cdot)$ with $m = 1,\ldots, M$, to a modified version of the data at each iteration. More precisely, if $f(\cdot)$ is a classifier that assigns input x to one of the classes $\{1,\ldots,K\}$, the error rate of the training sample is given by:

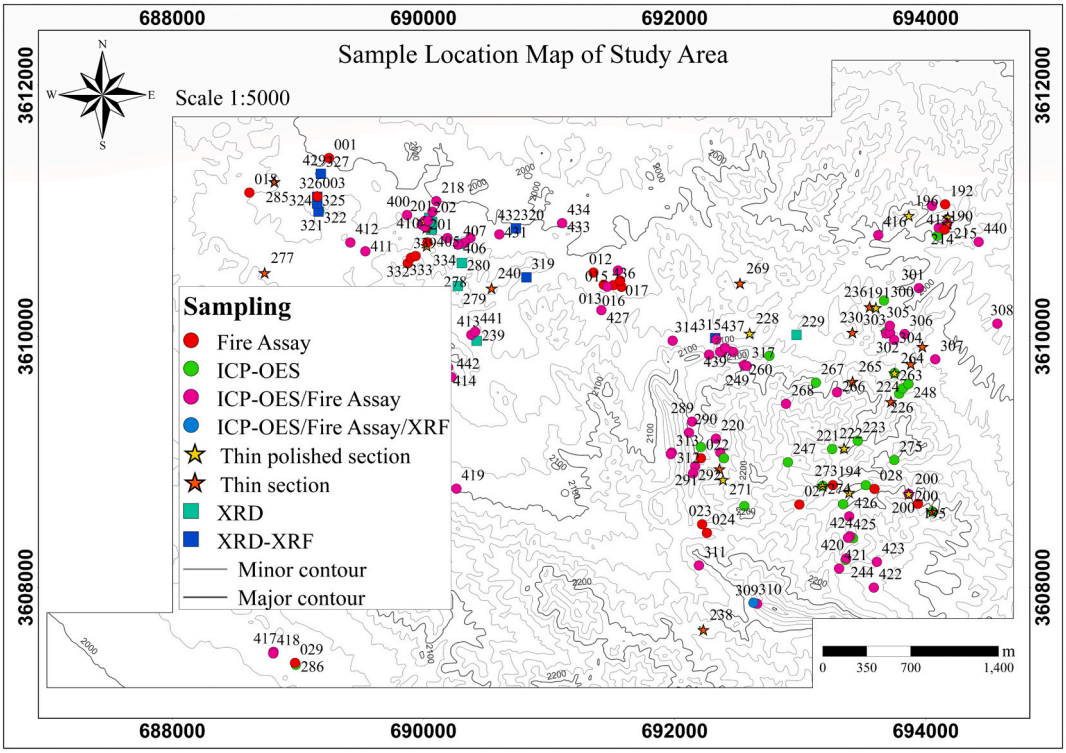

Fig. 4. Geographical distribution of sample locations in the study area.

$$\bar{\text{err}} = \frac{1}{N}\sum_{i=1}^{N} {}_1\{y_i \neq f(x_i)\} \tag{1}$$

where y_i is the true class of the input x_i. Moreover, the expected error rate on future predictions is given by $\mathbb{E}_{XY1}\{Y \neq f(X)\}$.

The final prediction in boosting is obtained by combining the predictions of the sequence of weak classifiers through a weighted sum or by taking the maximum of the weighted predictions:

$$f(x) = \max_m \{\alpha_m f_m(x)\} \tag{2}$$

where:

$$\alpha_m = log\left(\frac{1 - err_m}{err_m}\right) \quad \text{for } m = 1, \ldots, M \tag{3}$$

The effect of the α's is to give a higher influence to the more accurate classifier in the sequence.

Furthermore, the training dataset $\{(x_i, y_i) \text{ for } i = 1, \ldots, N\}$ are modified at each step applying weights $w_1, \ldots, w_N$, where

$$w_i^j = \begin{cases} \frac{1}{N} & \text{for } j = 1 \\ w_i^{j-1} e^{\alpha_m 1\{y_i \neq f_m(x_i)\}} & \text{for } j = 2, \ldots, M \end{cases} \tag{4}$$

for each $i = 1, \ldots, N$, where j indicates the iteration. At each iteration, these weights give more importance to misclassified observations of the previous classifier, whereas the importance (weight) for those classified correctly is decreased. Therefore, at each successive iteration, the new classifier is forced to pay more attention to the misclassified observations obtained by the previous classifier in the sequence. A diagram showing the boosting procedure is presented in Fig. 5.

In this study, the XGBoost algorithm was employed to study lithological classification. XGBoost is a specific variant of boosting and belongs to the Gradient Boosting Machine (GBM) algorithms. GBM methods are boosting techniques where the training of each model depends on the previously trained models. In GBM, the ith learner learns from the residuals of the $(i-1)$th leaner; in this way, the successive learners are maximally correlated with the negative gradient of the loss function. In particular, for XGBoost, the base learner is a tree model, and the objective function is given by:

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k),$$

where l is the loss function, $\hat{y}_i$ is the prediction for the ith instance, while y_i is the actual value, and finally Ω is a penalization component for the complexity of the model, to avoid overfitting. In particular, Ω is called the regularization component and, in XGBoost, takes the form:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2,$$

where T is the number of leafs and w is the leaf weights.

Since the XGBoost model is a boosting model and hence trained in an additive manner, the loss function to be minimized in the tth iteration is defined as:

$$L^{(t)} = \sum_i = l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t). \tag{5}$$

Hence, as indicated by Eq. 5, in the context of the tth tree, the aim is to predict the residuals while simultaneously striving to maintain a level of simplicity in the tth model.

3.4. Stacking

Wolpert (1992) introduced the stacked generalization technique, a.k.a. stacking, an ensemble method that enables the comparison of different learning algorithms. Stacking represents the inception of what is known as a "*meta learner*" in the literature, offering an alternative to the conventional concept of voting. In boosting, like in bagging, a voting procedure can be imposed after all the base learners have classified an

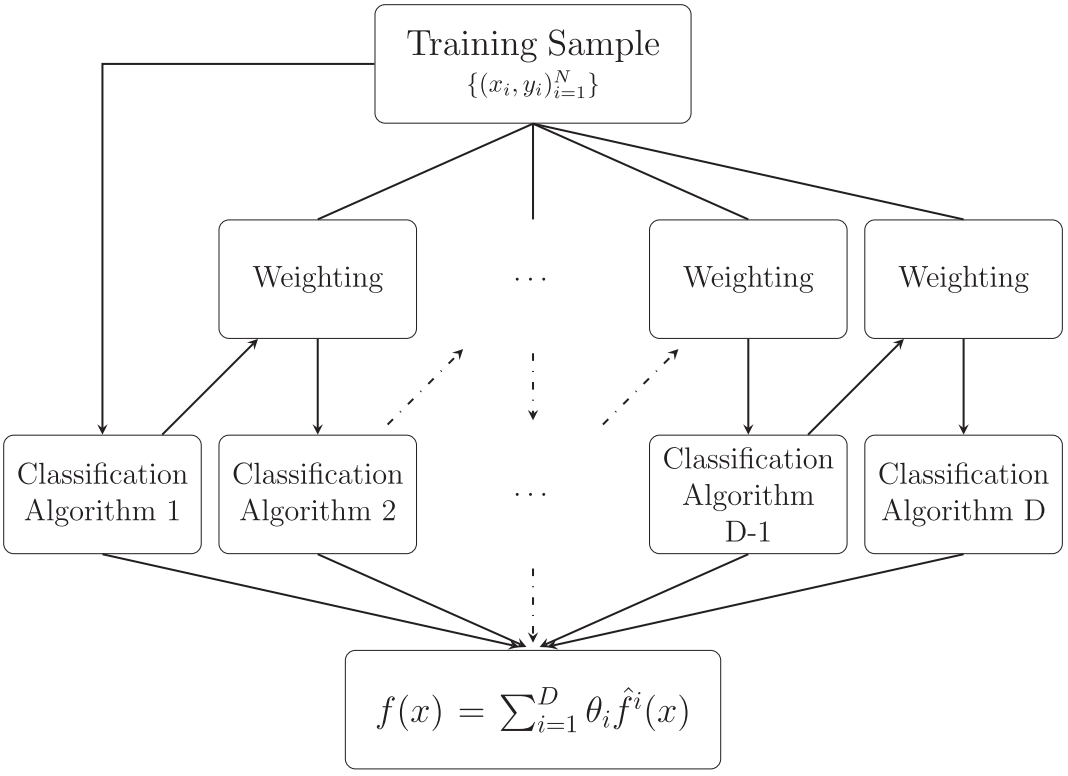

Fig. 5. Ensemble model: boosting.

instance, and the final class is determined based on their collective vote. Stacking, on the contrary, has two distinct learning layers. There is a *level-0 model* composed of the selected base models, which can vary, unlike in bagging and boosting, where base learners must be identical. Each base model can produce a different classification for a given instance. Instead of relying on a voting approach, the *level-1 model* combines these classifications (the base learners' outputs) to obtain a more accurate and robust classification.

In particular, the training dataset is split into two distinct training sets: one for the *level-0 model* and the other for the *level-1 model*. The initial training set is utilized to train the base models. At the end of this phase, the base learners classify the data in the second training set, which contains new data for all the base learners. For each instance within the second training set, a set of possible classes equal to the number of base learners in the ensemble is generated. Subsequently, the *level-1 model* is trained on these new data to learn how to combine the information to obtain a more accurate classification. A visual representation of the stacking procedure is illustrated in Fig. 6. It is worth noting that better results can be obtained using simple models for the *level-1 model*. For instance, prior studies, such as those by (Ting and Witten, 1999; Džeroski and Ženko, 2004) demonstrated improved performance by utilizing linear models and model trees as *level-1 models*).

StackingC represents a variant of stacking introduced by Seewald (2002) that aims to enhance the stacking performance in the context of multi-class classification problems. This modification is mainly built on Ting and Witten (1999), which uses a *multi-response linear regression* (MLR) model as *level-1 model*. In Ting and Witten (1999) the inputs of the MLR model were obtained by considering the class probabilities assigned by each base learner to the input instances. This information forms a table where rows represent data, and there are n sets of m columns, with n being the number of base learners and m representing the number of classes (Table 3 for a clarification). Each classifier provides probabilities for assigning the instance to each of the m classes. Seewald (2002), instead, gives as inputs to the MLR model m tables, one for each class, where only the class probabilities of the specified class were reported and the true belonging of the instance to the specified class were encoded as a 0–1 variable. Hence, Seewald trained the MLR model using m tables similar to Table 4. Obtaining better result when dealing with multi-class classification problems.

Table 3
Ting & Witten inputs for MLR.

$Classifier_1$			$Classifier_2$			…	$Classifier_n$		
c_1	…	c_m	c_1	…	c_m	…	c_1	…	c_m
$P_{1,c_1 1}$	…	$P_{2,c_m 1}$	$P_{2,c_1 1}$	…	$P_{n,c_m 1}$	…	$P_{n,c_1 1}$	…	$P_{n,c_m 1}$
$P_{1,c_1 2}$	…	$P_{2,c_m 2}$	$P_{2,c_1 2}$	…	$P_{n,c_m 2}$	…	$P_{n,c_1 2}$	…	$P_{n,c_m 2}$
⋮	⋮	⋮	⋮						
$P_{1,c_1 N}$	…	$P_{2,c_m N}$	$P_{2,c_1 N}$	…	$P_{n,c_m N}$	…	$P_{n,c_1 N}$	…	$P_{n,c_m N}$

3.5. Bagging

Bagging was introduced in Breiman (1996) and it is an ensemble method in which the building block is a given classifier or predictor. In this paper the chosen classifier is the decision tree, however, due to the hierarchical nature of classification trees, if data are added or changed,

Table 4
Seewald inputs for MLR.

$Classifier_1$	$Classifier_2$	…	$Classifier_n$	$Class$
c_i	c_i	…	c_i	$= = c_i$?
$P_{1,c_i 1}$	$P_{2,c_i 1}$	…	$P_{n,c_m 1}$	1
$P_{1,c_i 2}$	$P_{2,c_i 2}$	…	$P_{n,c_m 2}$	0
⋮	⋮	⋮	⋮	⋮
$P_{1,c_i N}$	$P_{2,c_i N}$	…	$P_{n,c_m N}$	1

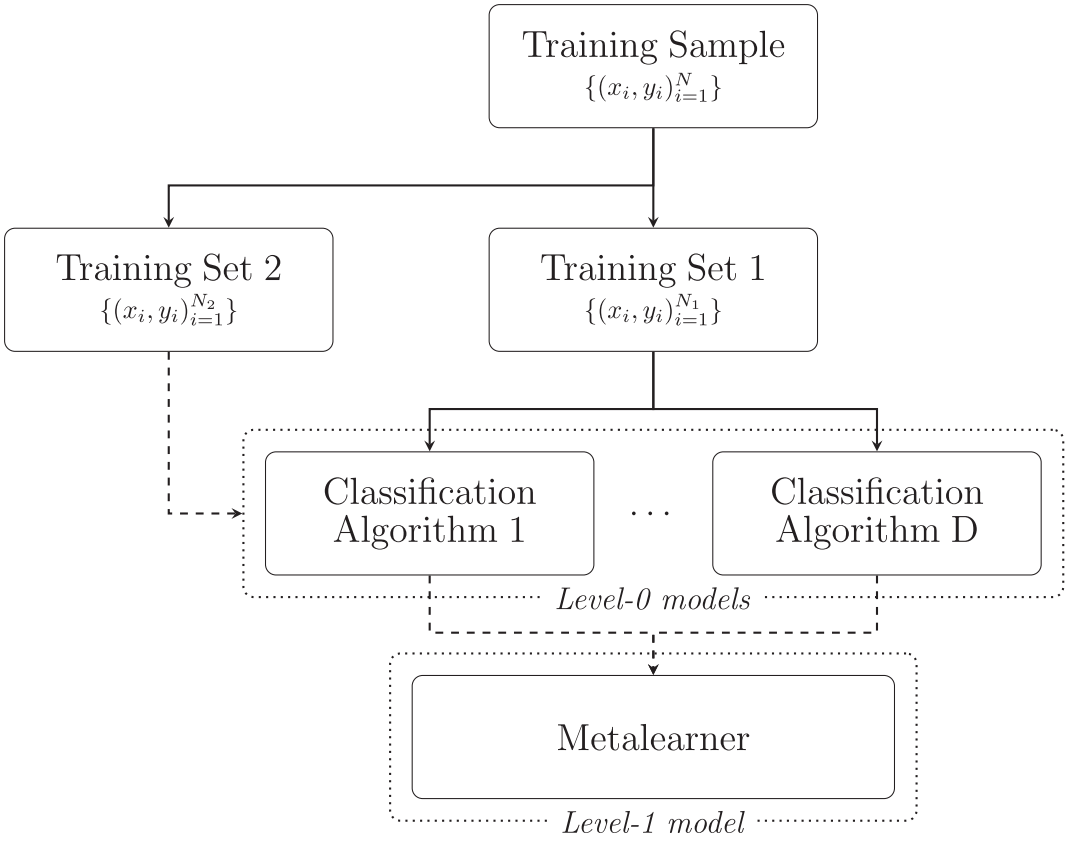

Fig. 6. Ensemble model: stacking.

these modifications will lead very easily to a different tree, therefore decision tree method suffers from high variance. Bagging is the solution to this problem, indeed Buhlmann and Yu (2002), Friedman and Hall (2007) showed that bagging presents smaller variance than simple decision trees.

Bagging uses the fact that given n independent observations $Z_1,\ldots,Z_n$, each with variance σ^2, then the variance of the mean $\overline{Z} = \sum_{i=1}^{n} Z_i$ is equal to $\frac{\sigma^2}{n}$. Hence, bagging reduces the variance by creating different decision trees, which have different classification rules, and then "averaging" the classification for the new data. More mathematically, in bagging a bootstrap is made on the training set (meaning that some groups of data are sampled from the dataset), in order to obtain D "new" datasets. Then, for each of these datasets, a decision tree is found, along with its classification rule, say $\widehat{f}^i(\cdot)$ for $i \in \{1,\ldots D\}$. Therefore, given a new data x, its predicted value is given by:

$$\widehat{f}_{\text{bag}}(x) = \frac{1}{D}\sum_{i=1}^{D}\widehat{f}^i(x) \tag{6}$$

In the case of classification trees, the above formula can be thought of as follows: the predicted class of x represents the most probable prediction among $\widehat{f}^i(\cdot)$ for $i \in \{1,\ldots D\}$, therefore, is like looking at the histogram of the predictions and then choose the prediction with more votes. A picture of bagging procedure scheme is presented in Fig. 7.

Random forest represents a variation of bagged trees. It was introduced independently by Ho (1995) and Amit and Geman (1994) and then developed by Breiman (2001). A random forest is a classifier consisting of a collection of tree-structured classifiers $\{f(x,\Theta_k),\ k = 1,\ldots\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors, and each tree casts a unit vote for the most popular class at input x. Rephrasing, this definition says that a random forest is a bagging tree where in each split of each tree, m of the p variables are randomly sampled, and then the split is allowed to use only one of those m predictors (usually $m \approx \sqrt{p}$).

Supposing there exists one powerful predictor in the data set, along with many other moderately strong predictors, then most or all the trees of a bagged tree will use this strong predictor in the top split, leading to bagged trees that look like each other. This leads to a situation where the bagged trees closely resemble each other, resulting in highly correlated predictions. Furthermore, averaging over highly correlated quantities does not reduce the variance as much as considering uncorrelated quantities from the beginning. Consequently, random forests can be perceived as a refinement of the bagged tree approach. Furthermore, following Breiman (2001), with an increasing number of trees, the generalized error, denoted as PE^*, converge to a certian value for nearly all sequences $\Theta_1,\ldots$.

$$\mathbb{P}_{X,Y}\left(\mathbb{P}_{\Theta}(f(X,\Theta) = Y) - \underset{j\neq Y}{max}\mathbb{P}_{\Theta}(f(X,\Theta) = j) < 0\right) \tag{7}$$

Looking more carefully inside the parenthesis we can see that we compute the probability, given a fixed data X and a fixed class Y, with respect to all the Θ that the suitable class for X is different from Y, and then we compute this probability over all X and Y, so we are computing a general error over all X, Y and Θ. This result explains why random forests do not overfit as more trees are added but produce a limiting value of the generalization error.

3.6. Evaluation metrics

To evaluate the performance of the proposed models, several measurement metrics were employed, including accuracy, precision, recall, specificity, F1-score, Kappa statistic, and Matthews Correlation Coefficient (MCC). The accuracy score measures the proportion of correct predictions among all predictions made by the model. Precision measures the proportion of true positives among all positive predictions made by the model. Recall measures the proportion of corrected clas-

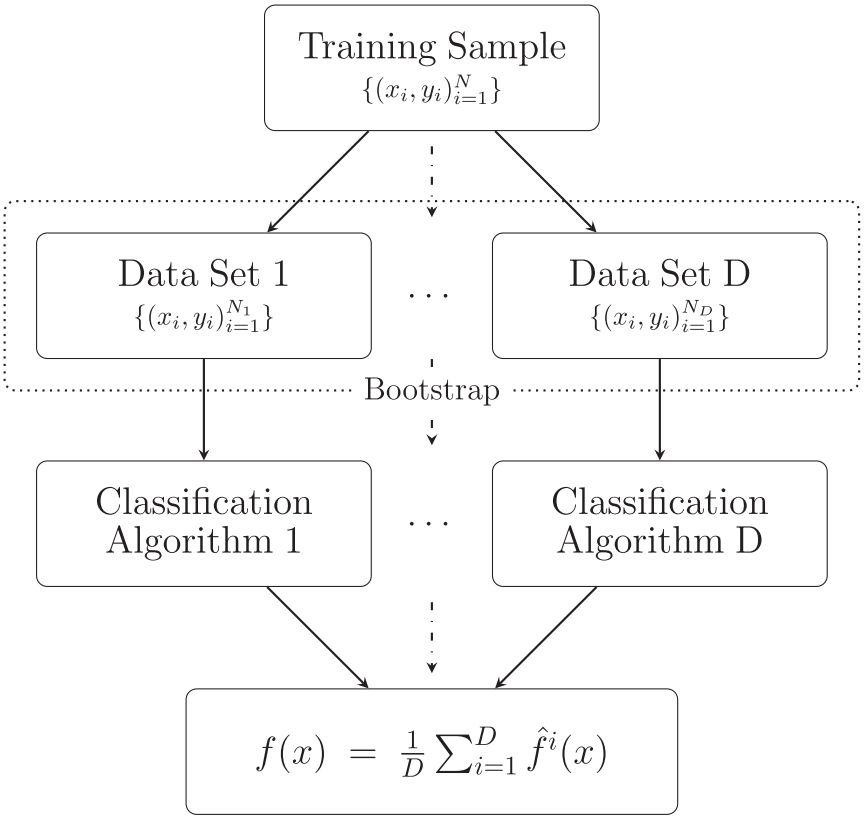

Fig. 7. Ensemble model: bagging.

sified data among all the possible data belonging to the same class; therefore, it measures in percentage how a class of data is entirely correctly classified. Specificity represents the proportion of true negatives among all negative classified data. F1-score combines Precision and Recall measures through a harmonic mean. Kappa statistic measures how closely the instances are classified correctly, controlling for the accuracy of a random classifier as measured by the expected accuracy. MCC measures the difference between the predicted classifications and the true ones in case of unbalanced classes. All the measures described above are presented below using formulae, where TP represents the number of true positives, TN represents the number of true negatives, FP represents the number of false positives, and FN represents the number of false negatives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$F1\text{-}score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

$$Cohen's\ Kappa\ (\kappa) = \frac{2(TP \cdot TN - FP \cdot FN)}{(TP + FP)(FP + TN) + (TP + FN)(FN + TN)}$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

4. Results and discussion

This section evaluates the performance of the proposed ensemble learning models by comparing them to two well-established benchmark models: LR and SVM. LR was first introduced in Cabrera (1994), while SVM was introduced in Vapnik and Chervonenkis (1974), and both models have since become foundational in the field of machine learning. Using these benchmark models as a reference point, the effectiveness of the proposed methods can be assessed. To compare the performance of the models, various measurement metrics are considered. Through this analysis, we can determine the superiority of the proposed ensemble models over the conventional models in predicting lithological classes.

The initial dataset for this study consisted of 280 samples collected from the study area, representing eight distinct lithological classes described by 12 features, including coordinates and elemental compositions resulting from geochemical analysis of rock samples. To ensure that subsequent analyses were based on a sufficiently large number of observations, classes with a sample size below 30 were removed. After applying this criterion, the final dataset contained 225 samples of four lithological classes. To avoid overfitting, the dataset was split into

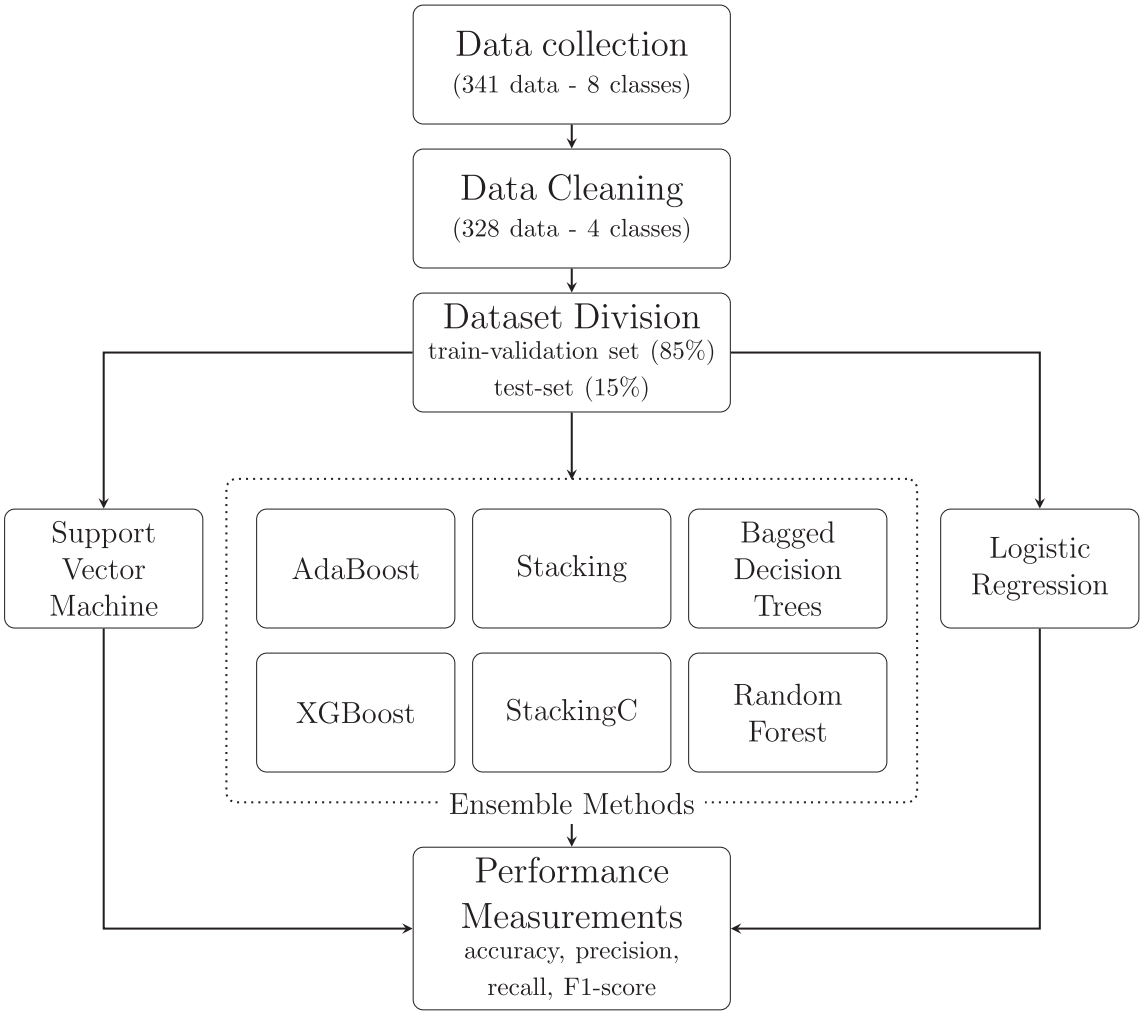

Fig. 8. The strategy workflow to analyze the proposed ensemble algorithms.

training and testing sets, accounting for 85 % and 15 % of the whole dataset, respectively. To further optimize model hyperparameters, a validation set was created, which accounted for 15 % of the training set (Fig. 8). These steps were taken to ensure the dataset's quality and consistency and improve the model's accuracy and generalization.

The label-encoder technique was used to transform categorical variables into numerical values. The primary motivation for using a label encoder is the limitation of most machine learning algorithms in handling categorical variables directly. The use of the label-encoder can lead to improvements in model performance and accuracy, particularly in classification tasks. The label-encoder has advantages for multi-class classification because it can represent each class as a unique numerical value. This process can simplify the computation of defined algorithms.

Boosting and bagging techniques employ decision tree models as their fundamental building blocks. This choice was made in order to align the base learner of these general techniques to the one used in the XGBoost and Random Forest algorithm. For this reason, in Table 5, instead of writing simply bagging, we prefer to specify that we used Bagged Decision Tree. Moreover, for a general boosting algorithm, the AdaBoost model is selected. For what concern stacking as *level-0 models* XGBoost and Random Forest were selected, which, as said before, are always based on decision trees, whereas an LR was applied as *level-1 model*. Ultimately, in stockings, extreme tree, random forest, and decision tree were implemented as *level-0* models, while MLR opted for *level-1* model. Furthermore, as explained in Section 3.4, the hole matrix of class probabilities was not used. Instead, submatrices were built and utilized to train the meta-learner to make the classification.

RandomizedSearchCV was applied to find the optimal hyperparameters in the proposed models. In this technique, the optimization of the hyperparameters is made randomly, meaning that for each hyperparameter, the algorithm will extract random values from a given probability density function. Given the non-finite nature of this algorithm, a maximum number of iterations should be imposed. Although RandomizedSearchCV usually provides more performing hyperparameters leading to better results (Bergstra and Bengio, 2012), it is usually time-consuming. The selected hyperparameters for each model using this technique are shown in Table 5.

Cross-validation is an important technique to optimize model performance and avoid overfitting in machine learning. In this study, we employed the repeated k-fold cross-validation method to evaluate the effectiveness of the proposed ensemble models in predicting lithological classes. As described in Ji-Hyun (2009), this method involves splitting the training set into k-folds, where each fold is considered a test set while the remaining $k - 1$ folds are used to train the model. This process is repeated multiple times, and the best-performing models are chosen.

Table 5
Hyperparameters for each model.

Model	Hyperparameter	Value
Logistic Regression	C	1.0
	penalty	elasticnet
	solver	lbfgs
Support Vector Machine	C	1
	kernel	rbf
	gamma	auto
Boosting	n_estimators	800
	learning_rate	0.8
	algorithm	SAMME
XGboost	n_estimators	800
	learning_rate	0.8
	max_depth	5.0
	subsample	1.0
Bagged Decision Trees	n_estimators	10
	max_samples	1.0
	max_features	0.8
Random Forest	n_estimators	200
	max_depth	10
	min_samples_split	gini

The repeated k-fold cross-validation method has several advantages over traditional methods such as leave-one-out or simple k-fold. One of the main advantages is its ability to provide a more accurate estimate of the model's performance, particularly for small datasets. This is because the repeated process reduces the high variability in model estimation in small datasets. By averaging the results obtained from multiple runs of the cross-validation process, the repeated k-fold cross-validation method allows for better model selection, which is essential for developing effective ensemble models. To assess the performance of each model, confusion matrices were generated (as shown in Fig. 10), which provide a summary of the classification model's effectiveness by comparing predicted and true labels of the test set. It is important to note that the confusion matrix as a visual representation of the model performance can provide a more detailed view of the model's performance compared to a single metric. The evaluation of the models was performed using various measurement metrics (3.6). These metrics enable a thorough assessment of each model's ability to predict the lithological classes, providing a comprehensive understanding of their performance.

Table 6 provides a comprehensive overview of the models' performance, assessing various performance metrics. Notably, all proposed ensemble models demonstrate remarkable accuracy, with scores ranging from 96.00 % to 98.00 %. The highest accuracy score, 98.00 %, is achieved by the AdaBoost, Bagged Decision Trees, and StackingC models. These results underline the effectiveness of these models in correctly classifying lithological samples. In contrast, the LR model shows a lower accuracy score at 62.00 %, which suggests that it might not be the most suitable choice for this specific geological classification task. Examining the precision and recall scores reveals valuable insights into the models' performance. The stackingC model achieves the highest precision score, 98.35 %, closely followed by the AdaBoost and Bagged Decision Trees models, both with 98.15 %. This highlights the stackingC model's capability to minimize false positives, making it a strong candidate for precise lithological classification. Specificity demonstrates that AdaBoost and Bagged Decision Trees with scores of 99.50 % outperform stackingC with scores of 98.80 %. This result suggests that these models excel in correctly identifying non-target lithological classes. The F1-score, a balance between precision and recall, is another critical indicator. StackingC achieves the highest F1-score of 98.10, which further emphasizes its balanced performance in minimizing both false positives and false negatives. In geological applications, this is particularly important as false positives can lead to significant errors in drilling and exploration operations. Additionally, Cohen's Kappa is an essential metric that measures the agreement between predicted and observed classifications. A higher Cohen's Kappa score indicates a stronger model. The stackingC, together with AdaBoost and Bagged Decision Trees models, with their high Cohen's Kappa score of 97.10 %, excels in achieving a substantial level of agreement between predictions and actual classifications, contributing to the reliability of the results. Moreover, the Matthews Correlation Coefficient (MCC) provides insights into the overall quality of the model's predictions. A higher MCC score suggests a more robust model performance. In this context, stackingC exhibits a significantly high MCC score of 93.70 %, emphasizing its strong predictive capability. These results demonstrate the superiority of the stackingC model in lithological classification, making it a compelling choice for geological mapping and resource exploration tasks, followed closely by Bagged Decision Trees and AdaBoost. For an in-depth understanding of the models' performance, radar charts were provided, as depicted in Fig. 9. These charts provide a holistic view of the models' effectiveness across utilized metrics.

Moreover, the confusion matrices were generated for each model to evaluate the models' performance in predicting the lithological classes (Fig. 10). These matrices present the predicted and actual labels of the test set and provide a detailed view of each model's performance. The Bagging and Boosting models had a high accuracy rate and correctly classified most samples in each class. On the other hand, the Random Forest model had a higher number of misclassifications for the K^{ar} unit

Table 6
Model performance comparison for predicting lithological classes– bolded values indicate top performance.

Model	Accuracy	Precision	Recall	Specificity	F1-score	Cohen's Kappa	MCC
Logistic Regression	62.00	51.60	51.60	91.40	54.20	37.70	43.00
Support Vector Machine	88.00	89.70	89.70	94.85	88.60	82.75	83.00
AdaBoost	**98.00**	98.15	98.15	**99.50**	97.95	**97.10**	92.80
Stacking	96.00	97.15	97.15	97.65	96.15	94.20	88.50
Bagged Decision Trees	**98.00**	98.15	98.15	**99.50**	97.95	**97.10**	92.80
XGBoost	96.00	97.15	97.15	97.65	96.15	94.20	88.50
StackingC	**98.00**	**98.35**	**98.35**	98.80	**98.10**	**97.10**	**93.70**
Random Forest	92.00	92.00	92.00	96.60	92.00	88.30	71.90

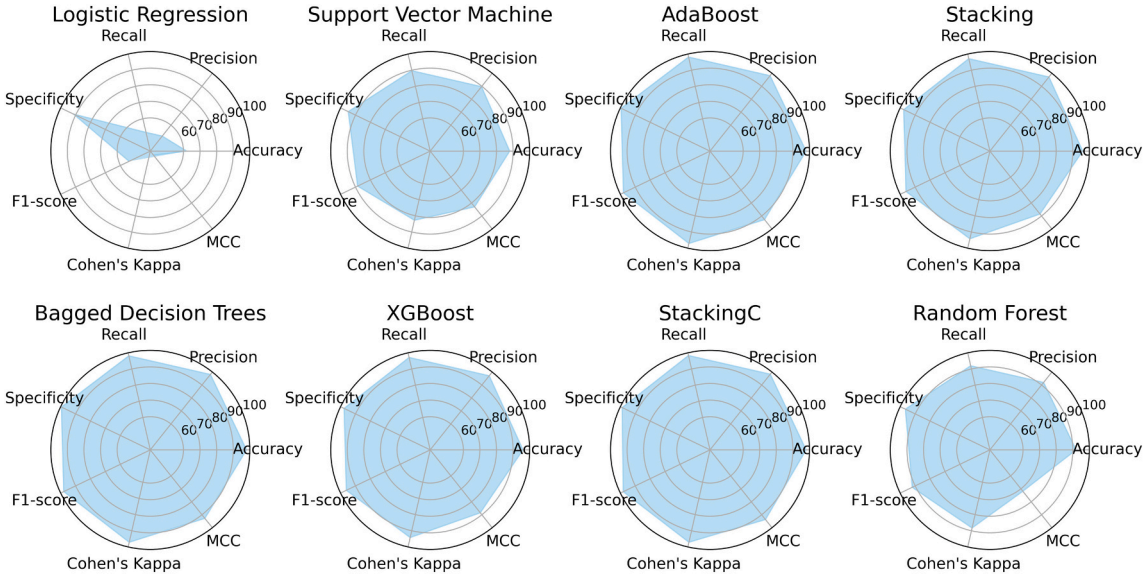

Fig. 9. Comparative radar charts showcasing the performance of different models, highlighting strengths and weaknesses across diverse evaluation metrics.

among the ensemble models. The LR and SVM models had the lowest overall performance, with many misclassifications for all classes.

The results of this study underscore the remarkable potential of ensemble learning techniques to significantly enhance lithological classification accuracy in the context of mineral exploration. Notably, stackingC is a top-performing model, surpassing the performances of the other introduced models across all evaluation metrics, positioning it as the most valuable model for this specific task. This is also highlighted by the agreement between stackingC class predictions and ground truth data in Fig. 11. Ensemble models are particularly effective at handling imbalanced datasets, such as the one used in this study, where some classes are much more dominant than others. Their impressive performance can be attributed to their capacity to aggregate predictions from multiple weak classifiers, thus reducing variance and enhancing the model's ability to generalize. Furthermore, these results also highlight the importance of choosing appropriate ML techniques for the specific task. Conventional methods like LR and SVM, commonly applied in classification tasks, exhibited poor performance in this study. This observation suggests that when dealing with intricate, high-dimensional data encountered in mineral exploration, more sophisticated models like ensemble learning techniques are better suited to yield accurate and reliable results.

Nevertheless, it is essential to clarify that these methods have some limitations. Boosting techniques, such as AdaBoost and XGBoost, while offering improved accuracy and robustness, can be sensitive to noisy data and require careful hyperparameter tuning to prevent overfitting. Stacking, which combines the strengths of various models, can introduce increased computational complexity and become time-consuming due to the need to train, store, and combine multiple models. Likewise, Bagging, though beneficial for reducing variance, may also lead to higher computational demands. Furthermore, the interpretability of the ensemble models is more complicated than that of the conventional models with just one learner. Therefore, even if the ensemble learner performs better, it causes a loss of the physical interpretability of the classification. Despite these potential limitations, it is crucial to emphasize that the advantages of employing these ML techniques in lithological classification significantly outweigh their drawbacks. The key to a successful application lies in the diligent selection and fine-tuning of these models, taking into account the specific requirements and characteristics of the dataset.

The findings presented in Fig. 11 offer a visual representation of the comparison between predicted and actual lithology classes utilizing conventional techniques, including LR, SVM, and best-performing model stacking techniques, emphasizing the critical role of accurate lithological classification in mineral exploration. The implication of these models extends to identifying potential mineralization in areas that traditional exploration methods may overlook. Nevertheless, it is essential to acknowledge that the performance of these models may vary based on each exploration site's unique geological and environmental conditions. Subsequent studies are imperative to comprehensively evaluate the effectiveness of these models in diverse contexts and determine their suitability for specific tasks. Beyond simply specifying potential mineral deposits, accurate classification enhances exploration efficiency, guides resource distribution, and reduces exploration costs. The adoption of ensemble learning techniques, such as Boosting, Stacking, and Bagging, significantly improves classification accuracy. However, a closer examination of this enhancement is revealed in the detailed performance metrics shown in Fig. 9. This figure visually presents metrics performance for each model, highlighting their strengths and areas for improvement. Notably, the stackingC model outperformed

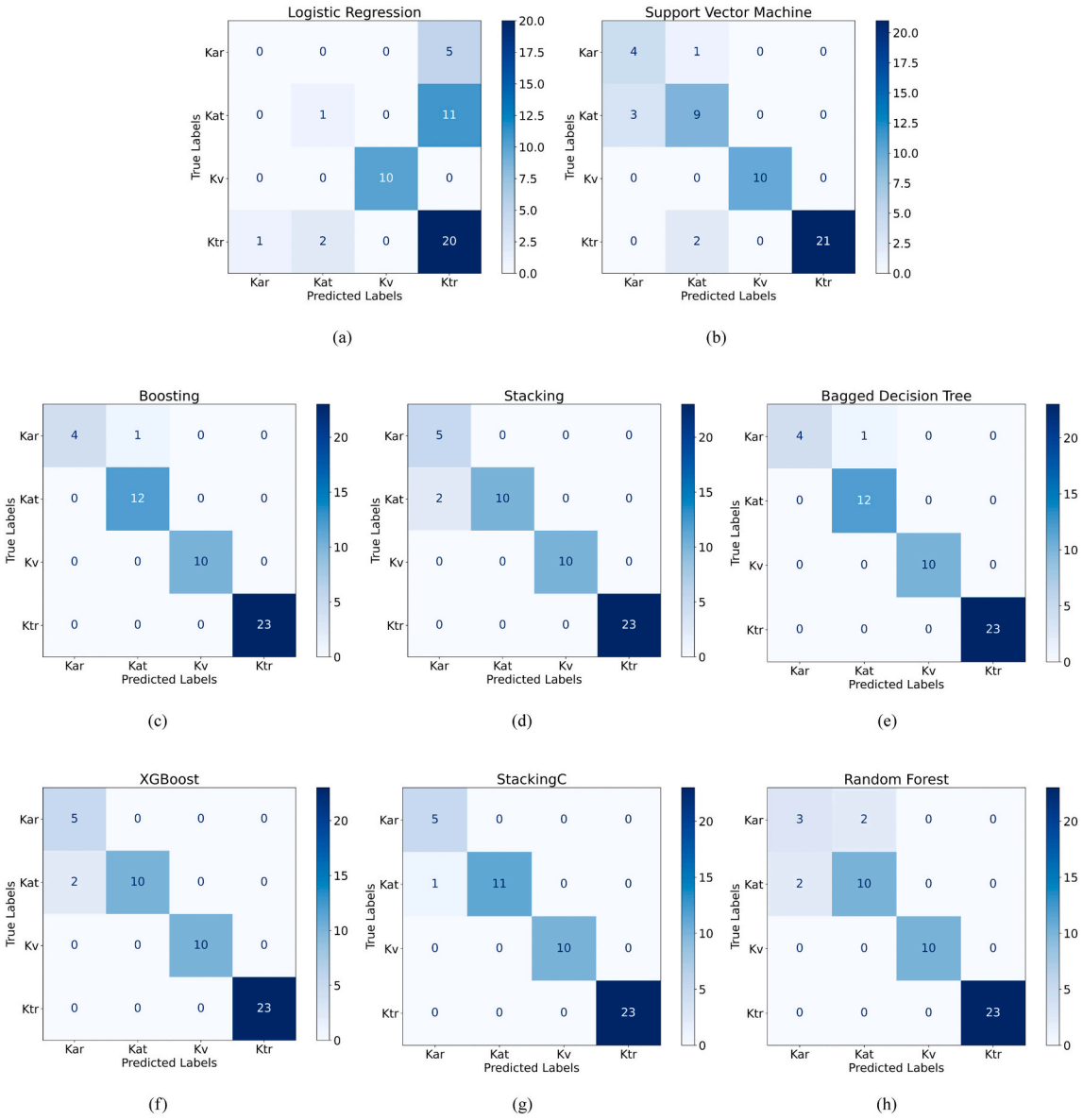

Fig. 10. Confusion matrices for the proposed model (a) Logistic Regression; (b) Support Vector Machine; (c) Boosting; (d) Stacking; (e) Bagging; (f) XGBoost; (g) StackingC (h) Random Forest.

all other models, showcasing a substantial improvement compared to conventional approaches. This comparison is important for gaining insights into each model's distinct contributions and refining strategies for future applications.

In Fig. 11, the locations of the samples collected from the study area during field studies for geochemical analysis and microscopic section preparation is illustrated. This spatial distribution of samples, completes the understanding of integrating the predicted lithology onto a geological map. These additional figures examine the models' outputs aligning with the exploration site's unique geological features. Notably, the regions highlighted by misclassifications, mainly those identified by the LR model, offer valuable insights into areas that may benefit from further model optimization. This detailed analysis provides a comprehensive perspective, enable to correlate misclassifications with specific geological conditions. The LR model shows room for improvement by misclassifying 19 units, highlighting an area for refinement. The SVM model outperforms with just 6 misclassified units, demonstrating enhanced performance. Impressively, the stackingC model aligns well with true lithology, except for a single exception marked on the map. This suggests that the stackingC model used in this study holds promise for enhancing geological maps. The overlap between predicted and actual lithology indicates its potential accuracy. The use of conventional approaches initially led to many misclassifications, but the subsequent adoption of ensemble models improved performance and results in this task. These findings underscore the effectiveness of ensemble learning techniques. Accurate geological maps, as demonstrated by these models, are crucial for successful geological and exploratory research.

5. Conclusions

In this study, the potential of ensemble learning models for lithological classification was extensively explored. The application of boosting, stacking, and bagging models demonstrated their ability to surpass conventional methods like LR and SVM. The findings underscore the valuable role of ensemble learning models in geoscience, with a specific emphasis on their importance in mineral exploration, where precise lithological classification is essential. Ensemble learning models exhibit a spectrum of advantages over traditional techniques, including

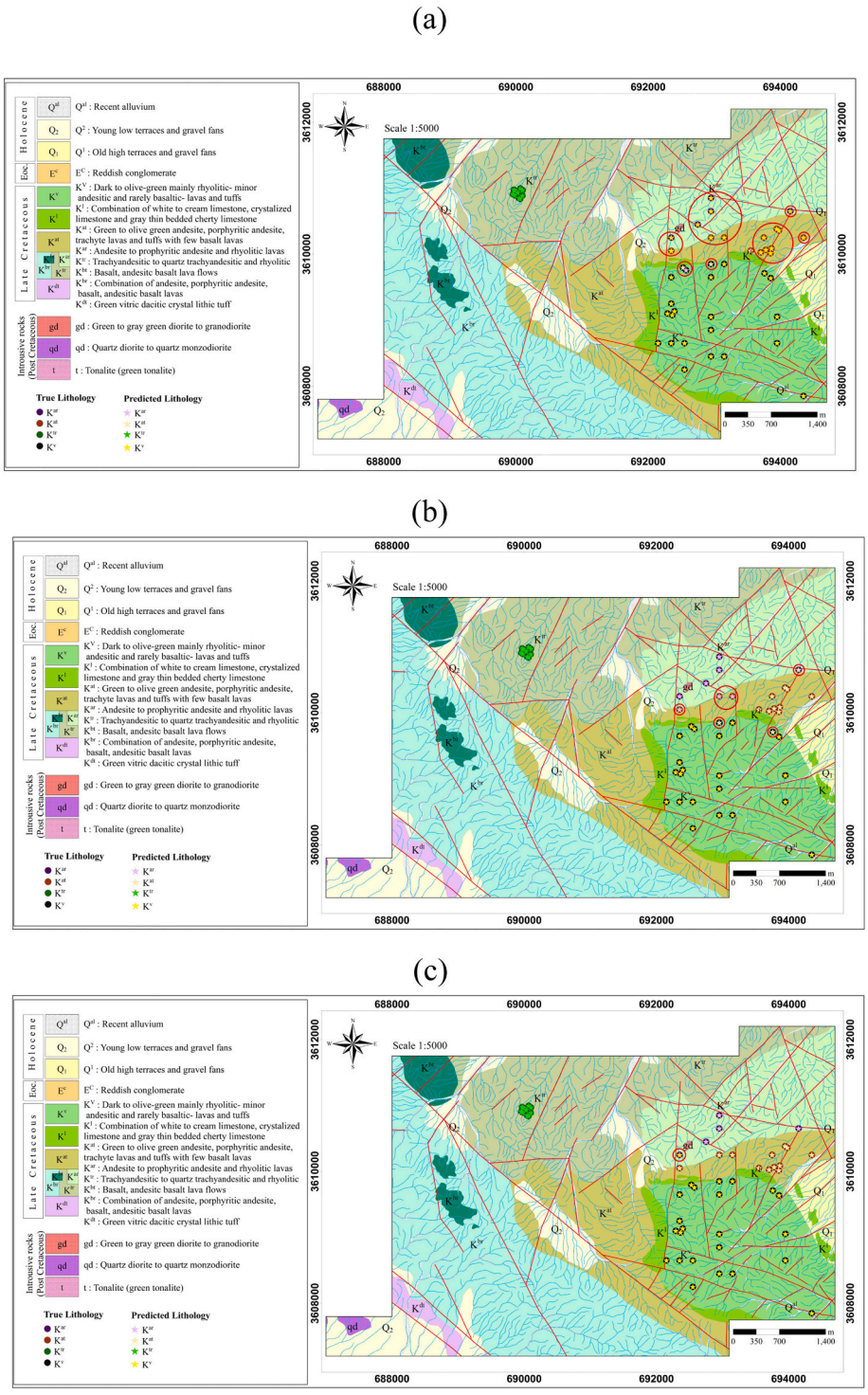

Fig. 11. Comparison of predicted and actual lithology classes: (a) linear regression, (b) support vector machine, (c) stackingC. The predicted lithology classes (triangles) are aligned with the actual lithology classes (squares) to assess their agreement. A single instance (indicated by the red circle) exhibits conflicting predictions between two classes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

their proficiency in mitigating overfitting, enhancing generalization, and improving model performance. Future research could explore the potential of ensemble learning models such as dynamic ensemble neural networks for broader applications. Additionally, more research and investigations can be conducted to refine these models, improve their generalization, and expand their scope of applications. It is important to emphasize that geological compositions may exhibit significant variability across different locations within an area. Therefore, the model's predictions can be validated through field observations and ground truth lithology.

CRediT authorship contribution statement

Sasan Farhadi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology. **Samuele Tatullo:** Writing – review & editing, Writing – original draft,

Visualization, Software, Methodology. **Mina Boveiri Konari:** Writing – original draft, Resources, Methodology, Investigation. **Peyman Afzal:** Validation, Supervision.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

Acknowledgment

The authors would like to thank Samaneh Kansar Zamin Company for providing valuable field information and analytical results.

References

Abedi, M., Norouzi, G.H., Bahroudi, A., 2012. Support vector machine for multi-classification of mineral prospectivity areas. Comput. Geosci. 46, 272–283.
Akkas, E., Akin, L., CubukCu, H.E., Artuner, H., 2015. Application of decision tree algorithm for classification and identification of natural minerals using SEM–EDS. Comput. Geosci. 80, 38–48.
Amit, Y., Geman, D., 1994. Randomized Inquiries About Shape: Application to Handwritten Digit Recognition. Chicago Univ IL Dept of Statistics.
Arganda-Carreras, I., Kaynig, V., Rueden, Eliceiri, K.W., Schindelin, J., Cardona, A., Seung, H.S., 2017. Trainable weka segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics 33, 2424–2426.
Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305.
Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Bressan, T.S., de Souza, M.K., Girelli, T.J., Chemale, F.J., 2020. Evaluation of machine learning methods for lithology classification using geophysical data. Comput. Geosci. 139.
Buhlmann, P., Yu, B., 2002. Analyzing bagging. Ann. Stat. 30, 927–961.
Cabrera, A., 1994. Logistic Regression Analysis in Higher Education: An Applied Perspective, vol. 10, pp. 225–256.
Choubin, B., Rahmati, O., 2021. Groundwater potential mapping using hybridization of simulated annealing and random forest. In: Water Engineering Modeling and Mathematic Tools.
Choubin, B., Hosseini, F., Rahmati, O., Youshanloei, M.M., Jalali, M., 2023. Mapping of salty aeolian dust-source potential areas: ensemble model or benchmark models? Sci. Total Environ. 877.
Das, I., Sahoo, S., van Westen, C., Stein, A., Hack, R., 2010. Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the Northern Himalayas (India). Geomorphology 114, 627–637.
Dercourt, J., Zonenshain, L., Ricou, L., Kazmin, V., Le Pichon, X., Knipper, A., Grandjacquet, C., Sbortshikov, I., Geyssant, J., Lepvrier, C., Pechersky, D., Boulin, J., Sibuet, J.-C., Savostin, L., Sorokhtin, O., Westphal, M., Bazhenov, M., Lauer, J., Biju-Duval, B., 1986. Geological evolution of the tethys belt from the Atlantic to the pamirs since the lias. Tectonophysics 123, 241–315.
Dev, V.A., Eden, M.R., 2019. Formation lithology classification using scalable gradient boosted decision trees. Comput. Chem. Eng. 128.
Dietterich, T., 2000. Ensemble methods in machine learning. In: multiple classifier systems. Lect. Notes Comput. Sci 1857, 3735–3745.
Džeroski, S., Ženko, B., 2004. Is combining classifiers with stacking better than selecting the best one? Mach. Learn. 54, 255–273.
Farhadi, S., Afzal, P., Boveiri Konari, M., Daneshvar Saein, L., Sadeghi, B., 2022. Combination of machine learning algorithms with concentration-area fractal method for soil geochemical anomaly detection in sediment-hosted Irankuh Pb-Zn deposit, central Iran. Minerals 12, 689.
Farhadi, S., Corrado, M., Borla, O., Ventura, G., 2024. Prestressing wire breakage monitoring using sound event detection. Comput. Aided Civ. Inf. Eng. 39(2), 186-202 39 (2), 186–202.
Felicisimo, A.M., Cuartero, A., Remondo, J., Quiros, E., 2013. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10, 175–189.
Friedman, J.H., Hall, P., 2007. On bagging and nonlinear estimation. J. Stat. Plann. Infer. 137, 669–683.
Gifford, C.M., Agah, A., 2010. Collaborative multi-agent rock facies classification from wireline well log data. Eng. Appl. Artif. Intel. 23.
Harris, D., Pan, G., 1999. Mineral favorability mapping: a comparison of artificial neural networks, logistic regression, and discriminant analysis. Nat. Resour. Res. 8, 93–109.
Harris, J.R., Grunsky, E.C., 2015. Predictive lithological mapping of canada's north using random forest classification applied to geophysical and geochemical data. Comput. Geosci. 80.
Hassanzadeh, J., Wernicke, B., 2022. The neotethyan Sanandaj-Sirjan zone of Iran as an archetype for passive margin-arc transitions. Tectonics 35.
Heidari, S., Afzal, P., Sadeghi, B., 2023. Miocene tectonic-magmatic events and gold/poly-metal mineralizations in the Takab-Delijan Belt, NW Iran. Geochemistry 83.
Ho, T.K., 1995. Random decision forests. In: IEEE Proceedings of 3rd International Conference on Document Analysis and Recognition, p. 1.
Ji-Hyun, K., 2009. Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. Comput. Stat. Data Anal. 53, 3735–3745.
Lima, R.P.D., Duarte, D., Nicholson, C., Marfurt, R., Slatt, K.J., 2020. Petrographic microfacies classification with deep convolutional neural networks. Comput. Geosci. 21423.
Liu, N., Xiaomei, L., Ershi, Q., Man, X., Ling, L., Gao, B., 2023. A novel ensemble learning paradigm for medical diagnosis with imbalanced data. IEEE Access 8, 171263–171280.
Maitre, J., Bouchard, K., Bedard, L.P., 2019. Mineral grains recognition using computer vision and machine learning. Comput. Geosci. 130, 84–93.
Mesaros, A., Heittola, T., Virtanen, T., Plumbley, M.D., 2021. Sound event detection: a tutorial. IEEE Signal Process. Mag. 38, 67–83.
Mosavi, A., Shirzadi, A., Choubin, B., Taromideh, F., Hosseini, F.S., Borji, M., Shahabi, H., Salvati, A., Dineva, A.A., 2020. Towards an ensemble machine learning model of random subspace based functional tree classifier for snow avalanche susceptibility mapping. IEEE Access 8, 145968–145983.
Movahednia, M., Maghfouri, S., Fazli, N., Rastad, E., Ghaderi, M., Gonzalez, F., 2022. Metallogeny of manto-type stratabound Cu-(Ag) mineralization in Iran: relationship with neotethyan evolution and implications for future exploration. Ore Geol. Rev. 149.
Ngiam, K.Y., Khor, I.W., 2019. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 20, e262–e273.
Otter, D.W., Medina, J.R., Kalita, J.K., 2021. A survey of the usages of deep learning for natural language processing. IEEE Trans. Neur. Netw. Learn. Syst. 32, 604–624.
Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M., 2015. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol. Rev. 71, 804–818.
Sagi, O., Rokach, L., 2018. Ensemble learning: a survey. WIREs Data Min. Knowl. Discov. 8.
Schapire, R., 1990. The strength of weak learnability. Mach. Learn. 5, 197–227.
Seewald, A., 2002. How to make stacking better and faster while also taking care of an unknown weakness. In: Proceedings of the 19th International Conference on Machine Learning, pp. 554–561.
Seo, H., Badiei Khuzani, M., Vasudevan, V., Huang, C., Ren, H., Xiao, R., Jia, X., Xing, L., 2020. Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications. Med. Phys. 47, e148–e167.
Stefenon, S., Corso, M., Nied, A., Perez, F., Yow, K., Gonzalez, G., Leithardt, V., 2022. Classification of insulators using neural network based on computer vision. IET Gener. Trans. Distrib. 16, 1096–1107.
Ting, K., Witten, I., 1999. Issues in stacked generalization. J Artif Intell Res 10.
Vapnik, V., Chervonenkis, A., 1974. Theory of Pattern Recognition.
Wang, J., Rao, C., Goh, M., Xiao, X., 2023. Risk assessment of coronary heart disease based on cloud-random forest. Artif. Intell. Rev. 56, 203–232.
Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A., 2020. Transformers: state-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics, pp. 38–45.
Wolpert, D.H., 1992. Stacked generalization. Neural Netw. 5, 241–259.
Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., Carson, P.L., 2019. Medical breast ultrasound image segmentation by machine learning. Ultrasonics 91, 1–9.
Yu, L., Porwal, A., Holden, E.J., Dentith, M.C., 2012. Towards automatic lithological classification from remote sensing data using support vector machines. Comput. Geosci. 45, 229–239.