Tree-Based Machine Learning Can Determine Lithofacies Properties of Reservoir Rocks- Camal Oil field, Yemen

Ghareb M. Hamada[1*], *Abbas M.* Al-khudafi[2], and *Hamzah A.* Al-Sharifi[2]

[1]1Arab Academy for Science, Engineering & Maritime Transport, Alexandria, Egypt
[2] Emirates International University, Sanaa, Yemen
[*]Corresponding author e-mail: ghareb.hamada@aast.edu

Abstract. This study aims to assess the effectiveness of several decision tree machine learning techniques for identifying formation lithology. A total of 20966 log data points from four wells were used to create the study's data. Lithology is determined using seven log parameters. The seven log parameters are the density log, neutron log, sonic log, gamma ray log, deep lateral log, shallow lateral log, and resistivity log. Different decision tree-based algorithms for classification approaches were applied. Several ML models, namely, the random forest, random decision tree algorithm, C4.5 decision tree, reduced-error pruning decision tree algorithm, logistic model trees, and Hoeffding tree, were assessed using well logging data for formation lithology prediction. The obtained results shows that the random forest model, out of the proposed decision tree models, performed best at lithology identification, with precession, recall, and F-score values of 0.913, 0.914, and 0.913, respectively. Random trees came next. with average precision, recall, and F1-score of 0.837, 0.84, and 0.837, respectively, the J48 model came in third place. The Hoeffding Tree classification model, however, showed the worst performance. We conclude that boosting strategies enhance the performance of tree-based models. Evaluation of prediction capability of models is also carried out using different datasets.

1 Introduction

Lithology can be established using well-log data to explore and produce petroleum. The lithology model of a reservoir can be established form other methods too, such as sedimentological classification on core data. Due to the intricacy of lithology, the distributions of logging data from distinct lithologies overlap, expanding the number of possible identifications. Thus, it is essential to use methods that provide an accurate means of forecasting lithology. Researchers have recently become more interested in applying machine learning approaches to forecast different types of lithology [27-32]. These approaches to lithology identification based on machine learning make an effort to train a multi-class classifier model based on a large amount of labelled well-logging data with logging curves, such as gamma ray (GR) resistivity logs, sonic logs, neutron logs, and density logs.

Various machine learning approaches have been proposed for the lithology classification problem. In lithological identification using logging data points, an artificial neural network first used to classify lithology [3,16]. Support Vector Machine (SVM) was utilized [1] to classify the lithology with logging data points and have accurately identified the lithology facies of heterogeneous sandstone reservoirs. Different types of multi-classification SVM were applied to identify volcanic lithology with well log data [6]. Random Forest was utilized to predict lithological mapping based on geophysical and geochemical data [9]. In the field of spatial modelling and classification based on well-log data, researchers are exploring various approaches to model rock lithology. This includes the development of novel hybrid inferential systems like ANN-HMM models, as well as the application of established techniques such as recurrent neural networks [2, 15].

Researchers proposed a groundbreaking method for real-time prediction of lithology during drilling operations. This approach leverages an artificial neural network (ANN) to integrate data from nearby wells with live drilling measurements from the South Pars gas field [13]. Using data from the Daniudui and Hangjinqi gas fields, five common machine learning techniques—Naïve Bayes, SVM, RF, Artificial Neural Network, and Gradient Tree Boosting—were assessed for detection of formation lithology [20].

Conventional single classification algorithms such as decision trees, SVM, and Bayes developed to determine the lithology of the Longqian region of China using well logs [8]. In order to predict the geological facies using well log data in the Anadarko Basin, Kansas, supervised learning algorithms, unsupervised learning algorithms, and a neural network machine learning algorithm were presented [12]. Generative adversarial networks were presented to recreate thin section images and identify carbonate lithology [14]. An Extreme Gradient Boosting and Bayesian Optimization classifier was proposed for identifying the lithology of the Daniudui and Hangjinqi gas fields [18]. The application of machine

* Corresponding author: ghareb.hamada@aast.edu

learning algorithms for real-time lithology determination during drilling operations was investigated by [17]. Three prominent algorithms are employed: neural networks (NNs), random forests (RFs), and extreme gradient boosting trees (XGBoosts). Additionally, a one-versus-one support vector machine (OVO SVM) approach is utilized for comparison. A coarse-to-fine architecture that incorporates outlier detection, multi-class classification, and a tree-based classifier suggested to identify the lithology using two actuals well logging data sets [19]. Chawshin et al. employed core CT scans in conjunction with convolutional neural networks (CNNs) for high-resolution lithology classification. Their study demonstrated that CNNs can accurately predict rock classes with high resolution, showcasing the effectiveness of this approach [32]. A hybrid framework consisting of artificial neural networks and hidden Markov models (ANN-HMM) was suggested for the classification of the lithological sequence [11]. They thoroughly evaluated the effectiveness of the suggested classifier using a combination of extreme gradient boosting (XGBoost) and Bayesian optimization (BO) [18]. Coal pay zones were predicted using a variety of machine learning algorithms (LR, SVM, ANN, RF, and XGBoost and data manipulation methods (NROS and SMOTE) [21]. Bi-directional gated recurrent units and a conditional random field layer (Bi-GRU-CRF) are the models used in the lithological sequence classification technique that was proposed by [10]. A hybrid framework consisting of artificial neural networks and hidden Markov models (ANN-HMM) was suggested for the classification of the lithological sequence Using the neural networks and hidden Markov models (ANN-HMM) hybrid framework [7]. The performance of the gradient boosting decision tree (GBDT) model, which was validated in comparison with the ANN, SVM, AdaBoost, and RF classifiers, was demonstrated by [22]. A Gray Wolf Optimization Algorithm (GWO-SVM)-based automatic identification system for lithology logging has been presented in [11]. The current study undertakes a comprehensive examination of the application of machine learning methods for lithology classification, with a specific emphasis on evaluating the performance of various tree-based models. This investigation involves a detailed comparison of different tree-based models, a crucial aspect of many machine learning studies. By integrating these techniques and approaches, the current work contributes to the ongoing research in machine learning and geology, enhancing the accuracy and efficiency of lithology classification models.

2 Material

2.1 Well log data the text

A total of 20966 log data points from four wells were collected from the Camal oil field to perform the evaluation, including seven logging parameters (density log (RHOB), neutron log (NPHI), sonic log (DT), gamma ray log (GR), deep latero log (LLD), shallow latero log (LLS), and resistivity log (RES) with corresponding depths. The classification will be based on a dataset comprised of well log data from four wells within the Camal oil field. These well logs have been meticulously annotated with lithology classifications based on direct observations of core samples and drilling cuttings. The range of the seven feature parameters are listed in Table 1. The evaluation was also conducted based on the three datasets based on used input parameters.

Table 1. Summary of Parameters Used for Lithology Classification.

Parameter	Max	Min	Stand.Dev	Mean
RES	1952.27	0.23	273.34	112.99
LLD	2064.76	0.23	63.72	29.74
LLS	2064.76	0.22	100.03	33.60
Depth	6100	520	1555	3421
GR	139.37	7.87	21.36	43.69
RHOB	2.95	1.94	0.18	2.28
NPHI	0.45	0.01	0.10	0.27
DT	141.76	38.71	17.87	91.54

2.2 Lithology classification

The lithology type and log curves parameters are shown in Figure 1. The well log in the Figure 1 appears to show a well that has penetrated a sequence of sandstone, limestone, dolomite, and shale formations.

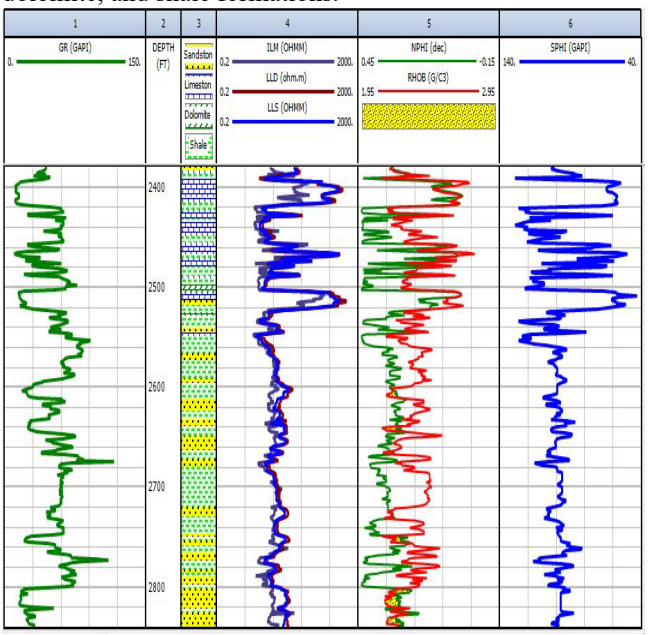

Fig. 1. The log curves and type of lithology of well 41 Camal Field.

The research aims to classify the lithology of subsurface formations, specifically identifying rock types such as shale, sand, sandstone, limestone, or dolomite. Figure 2 illustrates the interrelationships between various well log parameters.

Analysis of diverse parameter pairings reveals a distinct pattern of correlations. These correlations include robust relationships, such as the consistent association between LLD and LLS, and observable trends, such as the inverse relationship between GR and RHOB, which may be indicative of lithological influences. Overall, the graph suggests the presence of significant relationships between different well parameters, potentially offering valuable

insights into the geological composition of the logged formation.

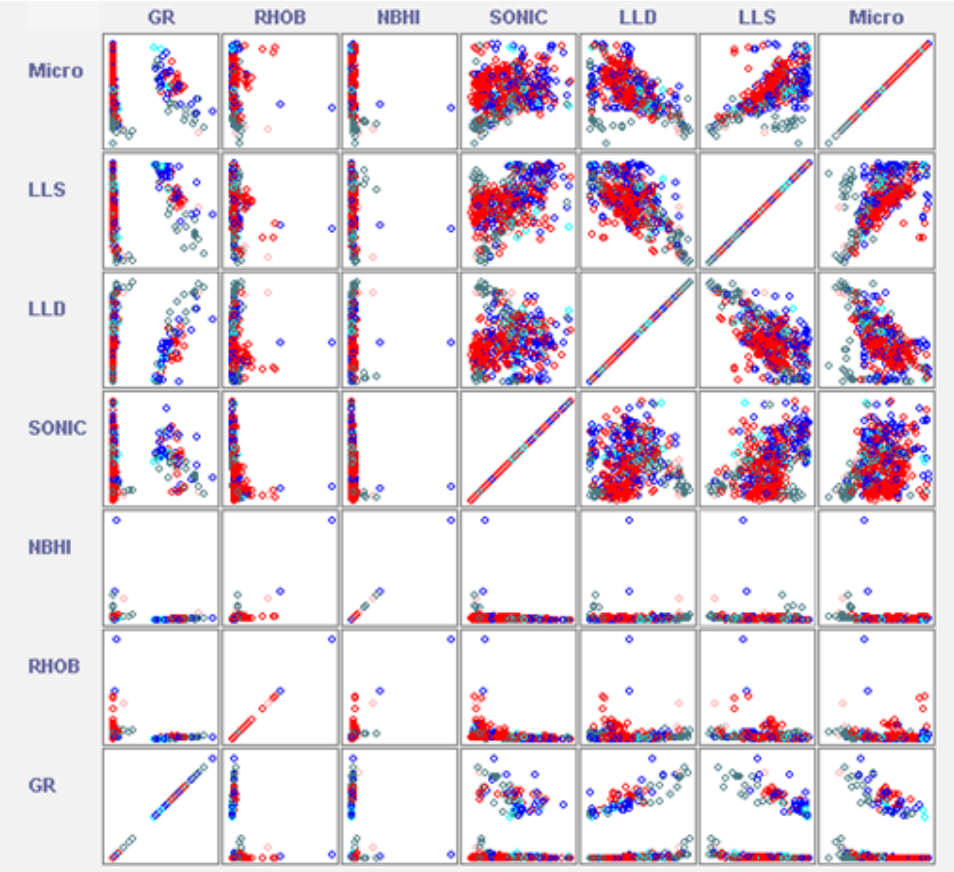

Fig. 2. Correlation of normalized log parameters.

3 Machine Learning Models

Different decision tree-based algorithms for classification approaches were applied. Six typical machine learning models, namely the Random Forest (RF). Random decision tree (RT), C4.5 decision tree (J48), reduced-error pruning decision trees (REPT), logistic model trees (LMT), and Hoeffding Tree (HT). Figure 3 presents the proposed lithology classification methods for this investigation.

3.1 Decision Tree

Three nodes make up a decision tree, which is a classification method: the leaf node, the branch (edge or link), and the root node. The test conditions for various attributes are represented by the root, all possible test outcomes are represented by the branch, and the labels of the classes to which the leaf nodes belong are present. The beginning of the tree, sometimes referred to as the top of the tree, is home to the root node. A decision tree is a hierarchical decision support model that uses a tree-like model of decisions and their potential repercussions, such as utility, resource costs, and chance event outcomes. It is one method of presenting an algorithm with just conditional control statements. In operations research, decision analysis in particular, decision trees are frequently utilized.

3.2 Random Forest

Known also as random decision forests, random forests are an ensemble learning technique that builds a large number of decision trees during the training phase for tasks like regression and classification. The class that the majority of the trees choose is the random forest's output for classification problems. The mean or average prediction made by each individual tree is returned for regression tasks [23]. The tendency of decision trees to overfit their training set is compensated for by random decision forests. Although they are less accurate than gradient-boosted trees, random forests still perform better than choice trees in most cases. Performance, however, might be impacted by data properties.

3.3 Reduced-Error Pruning Decision Tree

The Reduced-Error Pruning tree algorithm is a decision tree learning method with reduced-error pruning designed for efficient construction of classification or regression models. It builds the tree structure by employing information gain or variance for attribute selection at each node. Following the construction phase, REPT refines the tree using a technique called Reduced-Error Pruning (REP). This pruning step enhances the model's performance by removing subtrees that contribute to overfitting on the training data.

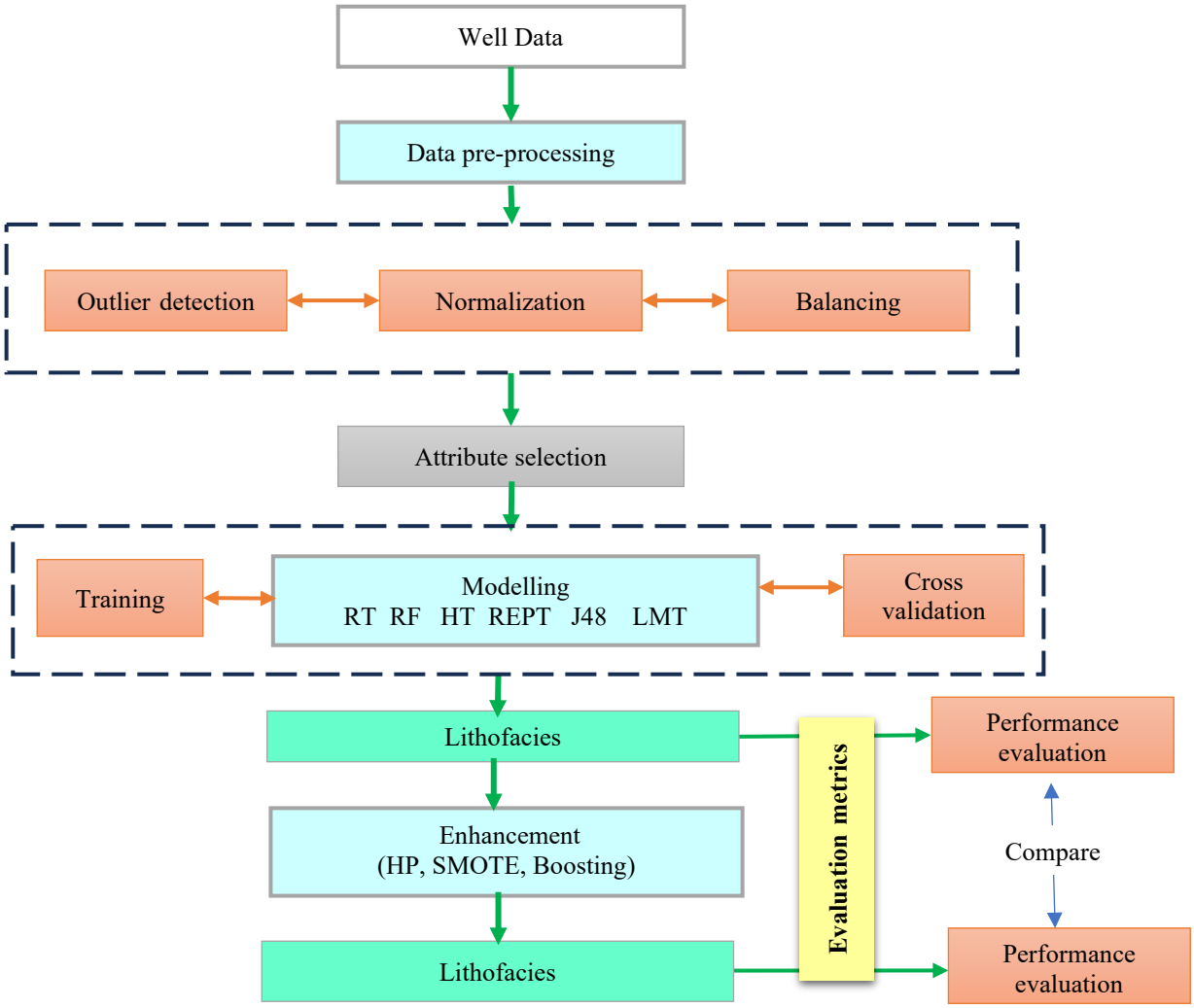

Fig. 3. Workflow for evaluation machine learning model.

3.4 Logistic Model Tree

Combining logistic regression and decision tree learning, the logistic model tree is a classification model that comes with a corresponding supervised training algorithm [24]. The concept of a logistic model tree is derived from the previous concept of a model tree, which is a decision tree with linear regression models at the leaves that generates a piecewise linear regression model instead of the piecewise constant model that would be produced by regular decision trees with constants at the leaves [24].

3.5 Hoeffding Tree

A Hoeffding tree, also known as Very Fast Decision Tree (VFDT), is a type of decision tree learning algorithm used for stream data classification. It's specifically designed to handle massive data streams, where data arrives continuously over time. Un-like traditional decision tree algorithms that require all the data to be available at once, Hoeffding trees can learn and update the model incrementally as new data points arrive.

3.6 C4.5 Classifier

The C4.5 algorithm, developed by Ross Quinlan, is a decision tree algorithm widely used for classification tasks. It is an ex-tension of the id3 algorithm and is often referred to as a statistical classifier. The C4.5 algorithm, an extension of J.R. Quinlan's earlier ID3 (Iterative Dichotomiser 3) model [4], builds decision trees through a recursive process. This entails progressively splitting the training data into subsets based on attributes that maximize information gain. In essence, C4.5 refines upon ID3 by addressing some of its inherent limitations [4].

4 Data Preprocessing

A dataset of 20,966 well log data points, encompassing Seven features (density log (RHOB), neutron log (NPHI), sonic log (DT), gamma ray log (GR), deep latero log

(LLD), shallow latero log (LLS)), and resistivity log (RES), was utilized for lithology classification.

4.1 Outlier Removal

This study employs the interquartile range (IQR) unsupervised learning technique, a robust outlier detection method, to identify and remove data points that may deviate significantly from the central tendency. This approach is implemented in conjunction with a filter-based classification algorithm. To assess the influence of outliers on the analysis, the model's performance was evaluated with and without outlier removal. Our analysis of the current results suggests that outliers exert a minimal influence on the model's accuracy. Consequently, the inclusion of this process may not be essential for achieving optimal performance.

4.2 Manage imbalanced dataset

To address imbalanced data and prevent overfitting or underperformance, we applied the Synthetic Minority Over-Sampling Technique (SMOTE). By increasing the proportion of minority instances in the dataset, this technique-maintained balance and enhanced algorithm performance. We employed the SMOTE function [5] specifically to tackle class imbalance issues related to different lithology types, enhancing lithology prediction model performance. The application of the SMOTE method improved the model's performance. For the random forest model, for instance, oversampling raised accuracy from 88.2% to 92.1%.

4.3 Normalization

Since logging indicators have varying dimensions, we performed data normalization after data collection, mining, and quality control. This step ensures consistency and allows us to combine dimensionless data to create new analysis indicators. All of the dataset's numerical values were standardized to fall between 0 and 1 before the machine learning model was trained.

5 Building Predictive Models

Tree-based models were constructed, namely RF, RT, J48, REPT, LMT, and HT. The dataset was meticulously divided into two mutually exclusive subsets : a training set and a testing set. The training set, typically comprising 80% of the data, serves as the foundation for model training. The learning algorithms utilize the training data to identify patterns and relationships within the features that predict the target variable. During the training phase, the chosen tree-based algorithms were exposed to the training set. They iteratively segmented the data based on specific features (independent variables) to create a tree-like structure. This structure essentially represents a series of decision rules that map the features to the target variable (dependent variable). Once trained, the models were evaluated using the unseen testing set (20% of the data). The testing set plays a crucial role in assessing the model's generalizability and ability to predict accurately on data it has not encountered during training. The classification models were also constructed using a ten-fold cross-validation technique.

5.1 Hyperparameters

Hyperparameters are parameters that control the learning process in machine learning models. Unlike other parameters, such as node weights, which are learned during training, hyperparameters are set beforehand [26]. They can be categorized as model hyperparameters, which influence model selection, or algorithm hyperparameters, which affect the learning process's speed and quality. Different machine learning algorithms require specific hyperparameters, and tuning them is crucial for adapting models to specific datasets [25]. Decision tree models leverage a series of hyperparameters to govern their complexity and, consequently, their performance. The primary hyperparameters influencing this complexity include: maximum depth, minimum samples per split, minimum samples per leaf, maximum features. Tuning these hyperparameters is crucial for striking a balance between model complexity and generalization performance. Techniques such as grid search or random search can be employed to efficiently explore the hyperparameter space and identify the optimal configuration for a given dataset.

5.2 Tuning Hyperparameter

To ensure optimal performance for lithology identification, we employed hyperparameter tuning for our machine learning models. A ten-fold cross-validation (CV) strategy was utilized to identify the most effective hyperparameter configuration for the tree model. Ten-fold CV is a well-established technique due to its ability to balance robust evaluation with computational efficiency. In this approach, the dataset is meticulously partitioned into ten equal-sized folds. The model is iteratively trained on nine folds and rigorously assessed on the remaining fold. This process is meticulously repeated ten times, guaranteeing that each fold is employed for validation once. The selection of ten folds represents a practical compromise: employing fewer folds increases the risk of overfitting, whereas a larger number becomes computationally expensive, approaching the computationally intensive leave-one-out CV method. By leveraging ten-fold CV, the model is exposed to a diverse range of data during training and testing, leading to a more reliable assessment of its ability to generalize to unseen data and significantly mitigating the risk of overfitting to the training data.

5.3 Gradient boosting approach

Boosting is a formidable ensemble learning technique in the realm of machine learning. It excels at elevating model performance by strategically combining multiple "weak learners" into a single, highly accurate "strong predictor." The core principle behind boosting lies in the sequential

training of these weak models. Each subsequent model meticulously focuses on the in-stances that were misclassified by the previous models. This iterative process involves adjusting the weights of the training examples, effectively forcing the following weak learners to concentrate on the challenging samples. By progressively refining the focus on these difficult instances, boosting algorithms gradually enhance the overall accuracy of the model. This process of aggregating the predictions from multiple weak models into a single, robust model has been demonstrably effective in significantly reducing both bias and variance. This makes boosting a highly potent approach for tackling complex machine learning challenges.

5.4 Attribute Selection

The correlation between the parameters may affect the ability of the model to forecast. Therefore, the analysis of input data is an important pre-processing technique beneficial for quality control, and through data mining, we can select more influen-tial parameters for lithology detection, reduce the dimension of the input parameter data, reduce overfitting, and improve accuracy. In this study, the importance of the features was evaluated for the prediction models. According to attribute rank, the major features contributing to lithology prediction were determined. This study employs several feature selection algorithms, including InfoGain, Relief, and OneR. To illustrate the feature selection process, we showcase the application of the Relief algorithm in detail. The results of this application are presented in Figure 4.

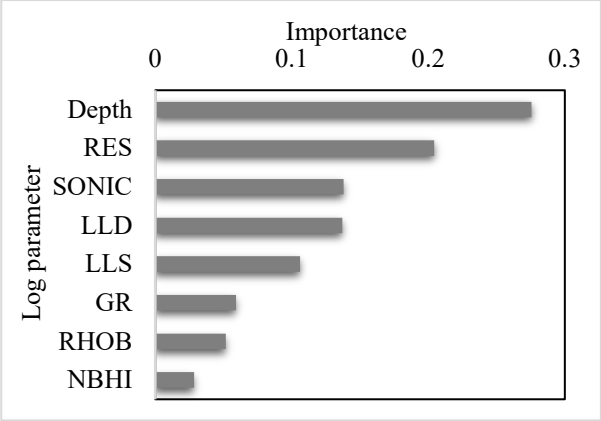

Fig.4. Importance rate of each feature for Relief algorithm.

Figure 4 presents a visual ranking of input variables based on their influence on lithology prediction using well log data. Depth emerges as the most significant feature, serving as a fundamental contextual element. Depth data impacts various aspects of analysis, including interpretation of well log responses, identification of formation boundaries, tracking of facies variations, and understanding of diagenetic effects. The inclusion of depth information demonstrably enhances the accuracy and efficiency of lithology classification within machine learning models. Table 2 presents datasets 1-8, which served as the foundation for a comprehensive evaluation of the models' prediction capabilities. This evaluation encompassed a wide variety of log parameter values to ensure the robustness of the models across different scenarios.

Table 2. Datasets of different functional forms.

Datasets	Log parameters
1	Depth, RHOB, GR, LLD
2	Depth, RHOB, GR, LLD, RES
3	Depth, LLD, ML
4	Depth, DT, LLD, LLS, RES
5	Depth, NBHI, RHOB, GR
6	Depth, NBHI, RHOB, GR, DT
7	Depth, RHOB, GR, LLS
8	Depth, RHOB, GR, LLS, RES

5.5 Model performance with different parameter sets

Figure 5 presents an analysis of how well logging features influence the performance of various modeling algorithms across different parameter sets. The evaluation metric employed is accuracy.

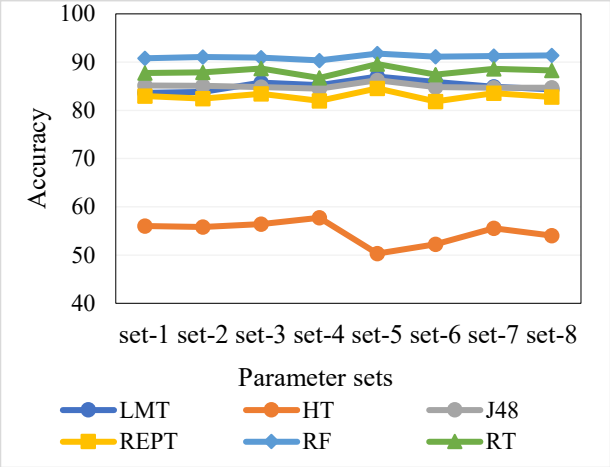

Fig.5. Performance of model with different parameter sets.

As shown from Figure 5 The LMT model exhibits consistent performance with accuracies ranging from 83.63% to 87.00%. This stability makes it a dependable option, although its accuracy falls within a moderate to high range compared to other models.

The HT model demonstrates higher variability in accuracy across the parameter sets, achieving a minimum of 50.31% and a maximum of 57.73%. This inconsistency suggests the need for further optimization to enhance its stability. The J48 model performs well, consistently maintaining accuracy levels above 84% across all sets. This consistent performance establishes its reliability for this dataset. The REPT model displays relatively stable performance with accuracies ranging from 81.99% to 84.55%. While demonstrating stability, its accuracy falls slightly below other models. The RF model stands out by consistently achieving the highest accuracy across all parameter sets, ranging from 90.36% to 91.76%. This exceptional performance makes it the most accurate and reliable model for this specific dataset. The RT model demonstrates a high level of accuracy, with values ranging from 86.72% to 89.60%. While exhibiting strong

performance, its accuracy remains slightly lower compared to the RF model.

The analysis reveals that the RF model surpasses all other models in terms of accuracy across all parameter sets. This exceptional performance establishes it as the most reliable choice for this dataset. J48 and LMT models also demonstrate strong and consistent performance, making them dependable alternatives. The HT model requires further optimization for improved consistency. These findings provide valuable insights for selecting the most suitable model based on the specific requirements of the dataset.

5.6 Model Performance and Parameter Sensitivity in Lithology Prediction

Figure 6 illustrates the relationship between the number of logging parameters utilized and the accuracy achieved by various lithology prediction models.

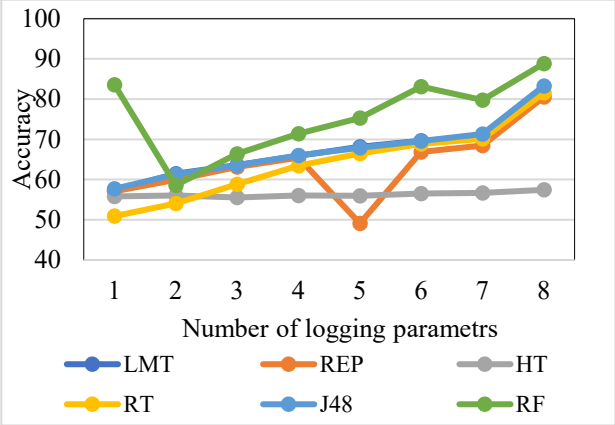

Fig.6. Influence of logging parameters on lithology prediction.

As shown from Figure 6 as the number of logging parameters increases, the accuracy of most models improves, with some exceptions. The RF model consistently outperforms the other models, achieving the highest accuracy across all parameter counts. With just one parameter, RF achieves an accuracy of 83.60%, which is significantly higher than the other models. As more parameters are added, the RF accuracy further increases, reaching 88.87% with eight parameters. The LMT, REP, and J48 models exhibit similar trends, with accuracy steadily rising as more parameters are included. LMT and J48 achieve comparable results, with LMT slightly outperforming J48 in most cases. The HT model demonstrates the lowest and most stable accuracy across all parameter counts, ranging from 55.54% to 57.50%. This suggests that HT is less sensitive to the number of logging parameters compared to the other models. The RT model shows a significant improvement as more parameters are added, reaching 81.57% with eight parameters. This indicates that RT benefits greatly from the inclusion of more logging parameters.

In conclusion, the RF model consistently outperforms the other models in lithology prediction, and its accuracy improves as more logging parameters are included. The LMT, REPT, and J48 models also show promising results, with accuracy increasing as more parameters are added. The HT model demonstrates the lowest and most stable accuracy, while the RT model exhibits a significant improvement in accuracy with the addition of more parameters.

6 Results and discussion

6.1 Evaluation of Machine Learning Algorithms

In this study, various evaluation metrics were used to assess the performance of classification models. These metrics included classification accuracy (AC), precision (Pr), recall (R), F-measure (F1), ROC area and the PRC area in order to more thoroughly assess the effectiveness of the learning model and the impact of lithology identification. Every classification model was assessed using ten-fold cross-validation. In Table 3, metrics scores for various models are presented.

Table 3. Average of evaluation metrics for different models-cross validation and training.

Model	Data	Pr	R	F1	ROC
RT	CV	0.895	0.896	0.895	0.927
	TR	0.901	0.902	0.901	0.935
RF	CV	0.919	0.92	0.919	0.988
	TR	0.913	0.914	0.913	0.985
REPT	CV	0.83	0.833	0.831	0.943
	TR	0.798	0.798	0.798	0.924
LMT	CV	0.835	0.836	0.835	0.927
	TR	0.833	0.833	0.833	0.928
J48	CV	0.848	0.85	0.848	0.897
	TR	0.837	0.84	0.837	0.891
HT	CV	0.452	0.552	0.427	0.614
	TR	0.525	0.538	0.516	0.737

Evaluation revealed that the Random Tree (RT) algorithm achieved strong performance on both datasets. Across all metrics (Precision, Recall, F1, and ROC AUC), RT scored above 0.89 for cross-validation and above 0.90 for training. While the Random Forest (RF) model also performed well, it fell slightly behind RT. RF secured F1 scores exceeding 0.91 and ROC AUC exceeding 0.98 on both datasets, with Precision and Recall mirroring this high performance.

Among the well-performing models, REPT exhibited the lowest scores. Its F1 scores hovered around 0.83, ROC AUC scores around 0.94, and both Precision and Recall remained near 0.83. The LMT demonstrated comparable performance to REPT, although achieving slightly better results on average. LMT's F1 score was approximately 0.83, ROC AUC around 0.93, and Precision and Recall scores both close to 0.83.

The J48 algorithm displayed lower performance compared to the previous three models. Its F1 scores averaged around 0.84, ROC AUC scores around 0.89, and both Precision and Recall scores settled at approximately 0.84. Finally, the Hoeffding Tree model exhibited the weakest performance overall. HT's F1 scores were around 0.43, ROC AUC scores around 0.61, and both Precision and Recall scores remained close to 0.5. The table suggests that the RT and RF models are the best performing models overall. They achieve the highest scores on all four metrics, and their performance is

consistent across both datasets. This suggests that they are not overfitting the training data. The remaining models (REPT, LMT, J48, and HT) all have lower performance. Overall, the table shows that RT and RF are the best performing models on this task.

Investigation of performance of tree models to identify different types of lithology is presented in Figure 7.

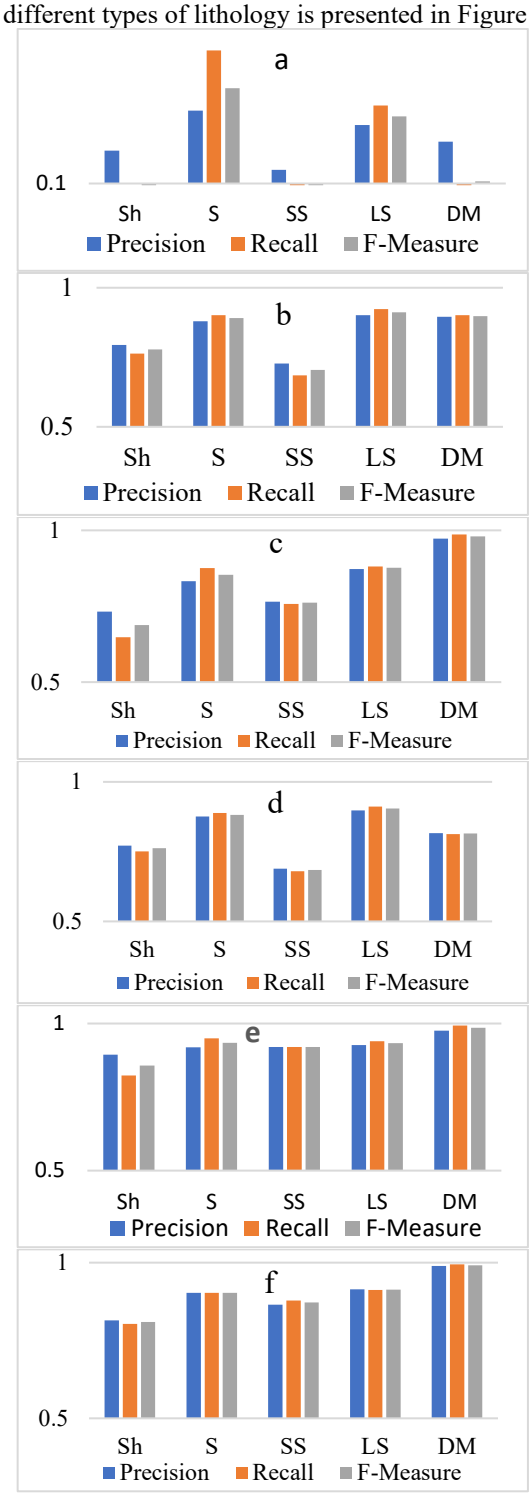

Fig. 7. Performance of tree models in identifying lithology classes: a) HT; b) J48; c) REPT; d) LMT; e) RF; f) RT

Analysis of the models' performance reveals that the RF model consistently achieves the highest precision, recall, and F-measure scores for both dolomite and limestone identification (Figure 7). This indicates that RF is the most accurate and reliable model for these tasks. For dolomite identification specifically, the RF model (Figure 7-e) outperforms all others with a precision of 0. 976, recall of 0. 993, and F-measure of 0. 985. While the REPT model (Figure 7-c) also exhibits strong performance with a very high recall (0. 986) and F-measure (0. 980), its precision (0. 973) falls slightly behind RF. The LMT model (Figure 7-d), on the other hand, demonstrates the lowest performance for dolomite identification, with all three metrics (precision: 0. 816, recall: 0. 813, F-measure: 0. 815) significantly lower than the other models. Similar to dolomite identification, the RF model reigns supreme for limestone with a precision of 0. 927, recall of 0. 940, and F-measure of 0. 933. Both the REPT and RT models display promising performance for limestone. REPT achieves a slightly higher precision (0. 872), while RT (figure 7-f) boasts a higher recall (0. 912) and F-measure (0. 913). Conversely, the HT model struggles significantly with lime-stone identification, exhibiting the lowest precision (0. 476), recall (0. 602), and F-measure (0.532) among all models. The RF model emerges as the clear leader for identifying both dolomite and limestone, consistently achieving the highest performance metrics. While the REP and RT models demonstrate strong capabilities, they fall short of RF's overall accuracy.HT model performs figure 7-a) the weakest, particularly for limestone identification, and the LMT exhibits the lowest performance for dolomite.

6.2 Confusion matrix

The confusion matrix was utilized to compare the performance of different models in classifying lithology classes. The confusion matrix presents the percentage of correctly classified instances for each lithology class. It highlighted instances where certain lithology classes were incorrectly identified as others. The confusion matrix of the lithologic classes, derived with an optimal technique, is shown in Table 4.

Table 4. Confusion matrix for different optimized classifiers using cross validation. a- j48, b- LMT, c- RT, d- RF, e- REPT

Actual class		Sh	S	SS	LS	DM
	Sh	22.36	3.33	0.34	0.62	0.02
	S	2.28	49.37	0.52	0.02	0.01
	SS	0.56	1.08	6.45	0.15	0.03
	LS	0.36	0.01	0.12	10.89	0.10
	DM	0.04	0.03	0.04	0.11	1.13
		Sh	S	SS	LS	DM

Predicted class (a)

Actual class		Sh	S	SS	LS	DM
	Sh	21.74	3.82	0.36	0.76	0.00
	S	2.88	48.68	0.62	0.01	0.01
	SS	0.64	1.27	6.12	0.19	0.05
	LS	0.52	0.03	0.15	10.73	0.07
	DM	0.04	0.02	0.06	0.14	1.09
		Sh	**S**	**SS**	**LS**	**DM**

Predicted class (b)

Actual class		Sh	S	SS	LS	DM
	Sh	17.76	4.77	0.94	0.81	0.07
	S	4.50	41.57	1.46	0.10	0.07
	SS	0.98	1.57	4.71	0.21	0.08
	LS	0.77	0.14	0.24	9.17	0.17
	DM	0.07	0.06	0.06	0.08	9.64
		Sh	**S**	**SS**	**LS**	**DM**

Predicted class (c)

Actual class						
	Sh	21.4	4.25	0.25	0.80	0.00
	S	1.95	49.93	0.31	0.00	0.01
	SS	0.82	1.76	5.45	0.21	0.02
	LS	0.25	0.07	0.09	11.00	0.09
	DM	0.07	0.04	0.03	0.10	1.11
		Sh	**S**	**SS**	**LS**	**DM**

Predicted class (d)

Actual class						
	Sh	19.35	3.96	0.31	0.73	0.00
	S	2.69	44.47	0.50	0.01	0.02
	SS	0.74	1.34	5.28	0.15	0.05
	LS	0.38	0.05	0.12	9.85	0.10
	DM	0.02	0.03	0.01	0.03	9.80
		Sh	**S**	**SS**	**LS**	**DM**

Predicted class (e)

6.3 Performance analysis of the Boosting algorithm

Boosting techniques, specifically AdaBoost meta-learners combined with classification tree models, were employed to enhance model performance.

As shown in the Figure 8, the boosting method yielded improved performance metrics compared to all other algorithms.

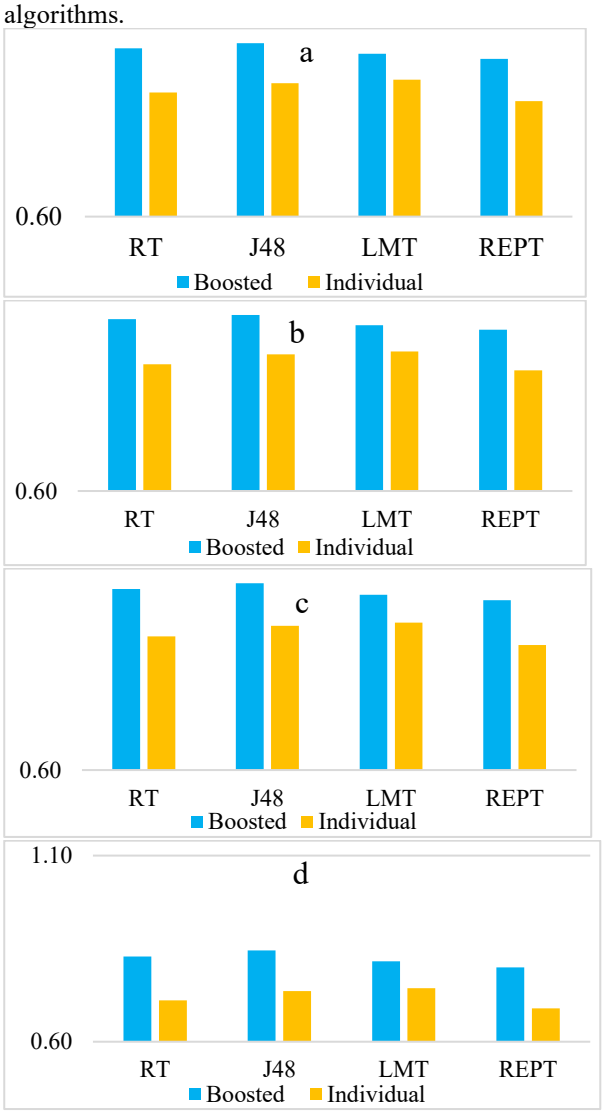

Fig. 8. Evaluation of individual and boosted model performance on: a) Precision) Recall) F-Measure; d) Matthews Correlation Coefficient.

The application of boosting significantly enhanced the performance of all machine learning models evaluated in this study. The most substantial improvement was observed in the Matthews Correlation Coefficient (MCC) metric. Boosting the REPT model yielded the largest increase in MCC (0.278), followed by the LMT model (0.128) and the RT model (0.12).

Furthermore, boosting led to modest improvements in precision (0.01 to 0.03) for all models except the LMT model, which experienced a slight decrease (0.01). Similarly, recall and F1-measure exhibited modest increases (0.01 to 0.03) for all models except the REPT model, where recall decreased slightly (0.005). Overall, boosting demonstrably improved the performance of all machine learning models on this task. The average improvement across all models was most evident in the MCC metric, with an increase of 0.176. It is crucial to acknowledge that the magnitude of improvement due to boosting can vary depending on the specific dataset and machine learning models employed. It is concluded that, ensemble methods such as boosting offer a powerful approach to enhancing the performance of machine learning models.

In figure 9, we evaluate the performance of individual and boosted models on various metrics, including precision, recall, F-measure, and the Matthews Correlation Coefficient.

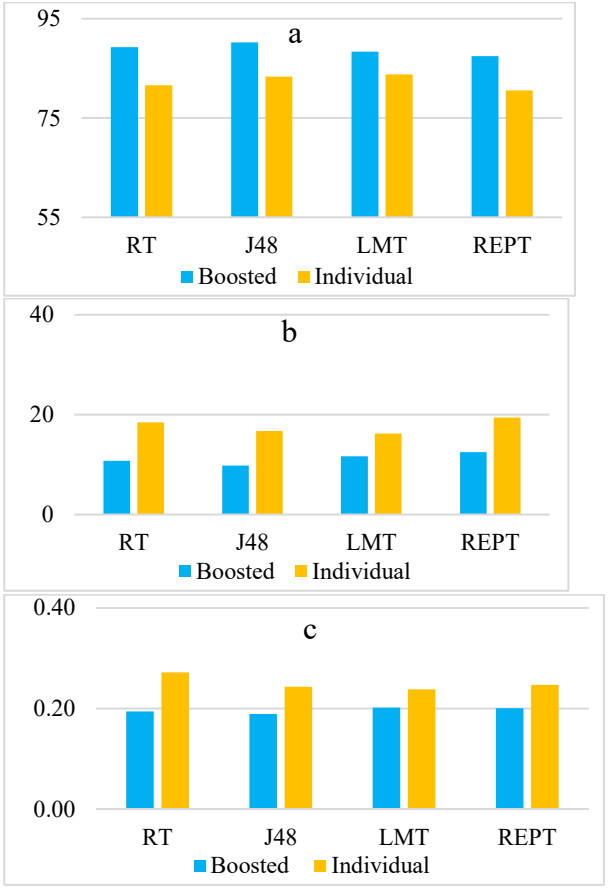

Fig. 9. Performance comparison of individual and boosted models. a) classification accuracy; b) classification error; c) RMSE; d) MAE.

Our findings reveal a consistent improvement in performance when employing boosted models compared

to their individual counterparts. This enhancement is evident across metrics that quantify both correctly classified instances and the magnitude of errors. For instance, the boosted RT model achieves an impressive 89.25% accuracy in classifying instances, surpassing the 81.57% achieved by the individual RT mod-el. This translates to a noteworthy improvement of 7.68%. Similarly, the boosted RT model exhibits a lower mean absolute error (MAE) of 0.0438 compared to the individual RT model's MAE of 0.0737, rep-resenting a reduction of 41.3%.

Furthermore, the kappa statistic, a measure of agreement between predicted and actual classes, demonstrates a positive impact from boosting. The boosted RT model boasts a kappa statistic of 0.8288, which is demonstrably higher than the 0.7093 achieved by the individual RT model. This translates to a significant improvement of 0.1195. The observed improvements in the kappa statistic ranged from 0.008 for J48 to 0.055 for LMT, highlighting the broad applicability of boosting across different model types.

The relative robustness and resistance to overfitting exhibited by boosted models are further corroborated by the relative absolute error and root relative squared error (RRSE) metrics. The boosted RT model demonstrates a superior performance with an RRSE of 54.3481, which is considerably lower than the 76.1253 observed for the individual RT model. This substantial reduction of 28.4% underscores the effectiveness of boosting in mitigating overfitting tendencies.

In conclusion, these results provide compelling evidence that boosting offers a powerful strategy for enhancing the performance of individual models. Boosted models consistently achieved superior accuracy, lower error rates, and improved agreement with actual classifications, demonstrating their effectiveness in this domain. The prediction capabilities of the models were evaluated using different datasets. Figure. 10 shows the prediction accuracy of different algorithms based on various datasets.

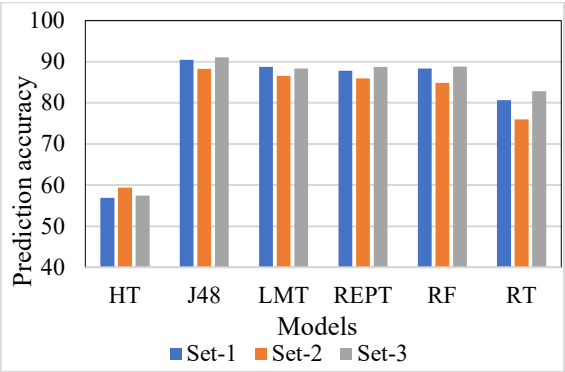

Fig. 10. Prediction performance of different models using various datasets.

The analysis reveals that the models exhibited comparable accuracy levels for set-3 and set-1. In particular, the J48 model demonstrated superior performance across both datasets, followed by the LMT model, while the HT model showed the lowest performance. The highest accuracy was achieved with set-3 across all models. Consistently, J48, LMT, REPT, and RF yielded similar results across various datasets. In contrast, the HT model consistently displayed the lowest accuracy across all three datasets.

6.4 Assessing the Generalizability of Tree-Based Models

In order to evaluate the model's ability to generalize to unseen data and assess its susceptibility to overfitting, the study employed three distinct datasets: training, cross-validation, and testing. The dataset was divided into three subsets for model development and evaluation. A total of 60% (12580 data points) of the data points were allocated for training the model. This allows the model to learn the underlying relationships within the data. An additional 20% (4193 data points) were designated for testing the model's performance on unseen data. This independent set helps assess the model's ability to generalize to new information. The remaining 20% (4193 data points) were used for cross-validation, a technique that iteratively splits the data for training and evaluation to ensure the robustness of the model's performance. The Figure 11 illustrates the performance comparaison of the Random Forest and J48 decision tree models across these three datasets. Figure 11 clearly illustrates that the Random Forest (RF) model achieves a perfect fit on the training data, with an accuracy of 100%.

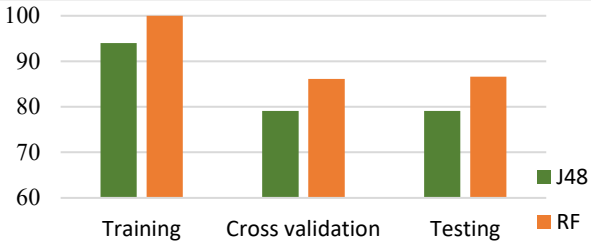

Fig.11. Investigating Model Behavior on a Range of Datasets

This is expected, given that the model is trained on this data and optimized to minimize the error. However, the cross-validation accuracy drops significantly to 86.14%, indicating that the model's performance does not generalize well to unseen data. The testing accuracy is marginally higher at 86.65%, indicating a slight improvement in the model's performance on unseen data. The RF model's performance is comprehensively evaluated across three datasets: training, cross-validation, and testing. While the model excels on the training data, it generalizes poorly to unseen data. The cross-validation and testing results collectively suggest that the model may be overfitting the training data and not generalizing well to new data. To enhance the model's performance, techniques such as regularization, feature selection, or ensemble methods could be employed to mitigate overfitting and improve generalization.

The J48 model exhibits a significant discrepancy between its performance on the training data (94 %) and both the cross-validation and testing datasets (79 %). This

stark difference suggests that the model may be overfitting the training data. Overfitting occurs when a model becomes overly attuned to the specific patterns and idiosyncrasies within the training data. While this leads to exceptional performance on the training set itself, the model fails to capture the underlying generalizable trends. Consequently, its ability to accurately classify new, unseen data suffers. In this case, the J48 model has meticulously learned the intricacies of the training data, resulting in high training accuracy. However, this success comes at the cost of generalizability. The model struggles to adapt to the variations present in the cross-validation and testing datasets, leading to a significant drop in accuracy. The J48 model's performance highlights the importance of considering generalizability when selecting a model.

6 Conclusions

This study investigated the effectiveness of machine learning models for lithology classification, employing a combination of data preprocessing, feature selection, hyperparameter tuning, and class imbalance mitigation techniques. Three well-logging datasets were utilized, and model evaluation was conducted rigorously through training, cross-validation, and further experimentation.

The study identified the crucial role of data preprocessing in improving classification accuracy. Feature selection algorithms like InfoGain, Relief, and OneR played a vital role in selecting informative parameters for lithology detection. This not only reduced overfitting but also enhanced prediction accuracy.

Hyperparameter tuning using a ten-fold cross-validation approach yielded optimal configurations, significantly improving the models' predictive capabilities. This is particularly evident in the case of the Random Forest model, where accuracy rose from 88.2% to 92.1% after applying SMOTE (Synthetic Minority Oversampling Technique) for class imbalance correction. This finding highlights the potential of combining SMOTE with machine learning algorithms for performance optimization.

Furthermore, the AdaBoost meta-learner, when combined with classification tree models, emerged as a highly effective approach. This boosting technique achieved superior performance metrics compared to individual models, reaching a remarkable accuracy of 89.25% in classifying lithological instances. This significantly surpassed the 81.57% accuracy achieved by the individual decision tree model (RT).

Overall, the study revealed that the Random Forest model outperformed other methods for lithology identification. This is further solidified by its precision (0.913), recall (0.914), and F-score (0.913) values. These findings demonstrate the efficacy of machine learning, particularly Random Forest models, for accurate lithology classification when combined with appropriate data manipulation and optimization techniques.

References

1. Al-Anazi, A., & Gates, I. D. On the capability of support vector machines to classify lithology from well logs. Natural Resources Research, **19**(2), 125–139. https://doi.org/10.1007/s11053-010-9118-9 (2010).
2. Alfarraj, M., & AlRegib, G. Petrophysical property estimation from seismic data using recurrent neural networks. SEG Technical Program Expanded Abstracts, 2141–2146. https://doi.org/10.1190/segam2018-2995752.1 (2018).
3. Al Ghathe, A.A., Hamada, G.M., and AlKhudifi, A.M., Water Saturation Determination in Carbonate Reservoirs Hybrid Artificial Intelligence Approaches and Conventional Techniques, MOC, April 19-21, Alexandria, Egypt (2016).
4. Alatefy S. Abdel Azim R., Alkouh A. and Hamada G., Integration of Multiple Bayesian Optimized Machine Learning Techniques and Conventional Well Logs for Accurate Prediction of Porosity in Carbonate Reservoirs, Processes (2023), **11**, 1339. https://doi.org/10.3390/pr11051339, (2023).
5. Chawla, N. V, Bowyer, K. W., & Hall, L. O. SMOTE: Synthetic Minority Over-sampling Technique. **16**, 321–357. (2002).
6. Dan, M., Zhu-Wen, W., Yu-Long, H., Shi, X., & Da-Peng, Z. Lithological identification of volcanic rocks from SVM well logging data: Case study in the eastern depression of Liaohe Basin. en.dzkx.org. https://doi.org/10.6038/cjg20150528 (2015).
7. Feng, R. Lithofacies classification based on a hybrid system of artificial neural networks and hidden Markov models. Geophysical Journal International, 221(3), 1484–1498. https://doi.org/10.1093/gji/ggaa083 (2020).
8. Gong, K., Ye, Z., Chen, D., Zhu, D., & Wang, W. Investigation on automatic recognition of stratigraphic lithology based on well logging data using ensemble learning algorithm. Society of Petroleum Engineers - SPE Asia Pacific Oil and Gas Conference and Exhibition 2018, APOGCE 2018, 1–11. https://doi.org/10.2118/192006-ms (2018).
9. Harris, J. R., & Grunsky, E. C. Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. Computers and Geosciences, 80, 9–25, https://doi.org/10.1016/j.cageo.2015.03.013, (2018).
10. Liu, Z., Cao, J., You, J., Chen, S., Lu, Y., & Zhou, P. A lithological sequence classification method with well log via SVM-assisted bi-directional GRU-CRF neural network. Journal of Petroleum Science and Engineering, 205(42030812), 108913. https://doi.org/10.1016/j.petrol.2021.108913 (2021)
11. Lu, S., Li, M., Luo, N., He, W., He, X., Gan, C., & Deng, R. Lithology Logging Recognition Technology Based on GWO-SVM Algorithm.

Mathematical Problems in Engineering, https://doi.org/10.1155/2022/1640096, (2022).

12. Mohamed, I. M., Mohamed, S., Mazher, I., & Chester, P. (2019). Formation lithology classification: Insights into machine learning methods. Proceedings - SPE Annual Technical Conference and Exhibition, (2019)-Septe. https://doi.org/10.2118/196096-ms.
13. Mohammad Ali, A. M. Artificial Intelligence for Lithology Identification through Real-Time Drilling Data. Journal of Earth Science & Climatic Change, 06(03), 3–6. https://doi.org/10.4172/2157-7617.1000265 (2015).
14. Nanjo, T., & Tanaka, S. Carbonate lithology identification wth generative adversarial networks. International Petroleum Technology Conference 2020, IPTC 2020. https://doi.org/10.2523/iptc-20226-ms (2020).
15. Pham, N., Wu, X., & Zabihi Naeini, E. (2020). Missing well log prediction using convolutional long short-term memory network. Geophysics, **85**(4), WA159–WA171. https://doi.org/10.1190/geo2019-0282.1.
16. Rogers, S. J., Fang, J. H., Karr, C. L., & Stanley, D. A. Determination of lithology from well logs using a neural network. In American Association of Petroleum Geologists Bulletin (Vol. **76**, Issue 5, pp. 731–739). https://doi.org/10.1306/bdff88bc-1718-11d7-8645000102c1865d (1992).
17. Sun, J., Chen, M., Li, Q., Ren, L., Dou, M., & Zhang, J. A new method for predicting formation lithology while drilling at horizontal well bit. Journal of Petroleum Science and Engineering, 196. https://doi.org/10.1016/j.petrol.2020.107955 (2021).
18. Sun, Z., Jiang, B., Li, X., Li, J., & Xiao, K. A data-driven approach for lithology identification based on parameter-optimized ensemble learning. Energies, **13**(15), 1–15. https://doi.org/10.3390/en13153903 (2020).
19. Xie, Y., Zhu, C., Hu, R., & Zhu, Z. A Coarse-to-Fine Approach for Intelligent Logging Lithology Identification with Extremely Randomized Trees. Mathematical Geosciences, **53** (5), 859–876. https://doi.org/10.1007/s11004-020-09885-y (2021).
20. Xie, Y., Zhu, C., Zhou, W., Li, Z., Liu, X., & Tu, M. 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. https://doi.org/10.1016/j.petrol.2017.10.028 (2018).
21. Zhong, R., Johnson, R. L., & Chen, Z. Using machine learning methods to identify coals from drilling and logging-while-drilling LWD data. SPE/AAPG/SEG Asia Pacific Unconventional Resources Technology Conference (2019), APUR (2019). https://doi.org/10.15530/ap-urtec-2019-198288 (2019).
22. Zou, Y., Chen, Y., & Deng, H. Gradient Boosting Decision Tree for Lithology Identification with Well Logs: A Case Study of Zhaoxian Gold Deposit, Shandong Peninsula, China. Natural Resources Research, **30** (5), 3197–3217. https://doi.org/10.1007/s11053-021-09894-6 (2021).
23. Ho, Kam, Tin. "The Random Subspace Method for Constructing Decision Forests." Ieee Transactions on Pattern Analysis and Machine Intelligence **20** (8):832–44. (1998).
24. Landwehr, Niels, Mark Hall, and Eibe Frank. "Logistic Model Trees." Machine Learning **59** (1–2):161–205. (2005).
25. Claesen, Marc, and Bart De Moor. (2015), MIC (2015): The XI Métaheuristiques International Conférence in Agadir, Morocco.
26. Yang, Li, and Abdallah Shami. "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice." Neurocomputing 415:295–316. (2020).
27. Fang, Zhengwei, Liqiang Zhang, and Shicui Yan. (2023). "Forecast of Lacustrine Shale Lithofacies Types in Continental Rift Basins Based on Machine Learning: A Case Study from Dongying Sag, Jiyang Depression, Bohai Bay Basin, China." Frontiers in Earth Science 11(April):1–15.
28. Lin, Xiaochun, and Shitao Yin. (2023). "Lithology Identification Based on Interpretability Integration Learning." Earth Science Informatics **16**(3):2211–22.
29. Potekhin, Denis V., and Sergei V. Galkin. (2023). "Use of Machine Learning Technology to Model the Distribution of Lithotypes in the Permo-Carboniferous Oil Deposit of the Usinskoye Field." Journal of Mining Institute 259:41–51.
30. Ziadat, Wael, Hany Gamal, and Salaheldin Elkatatny. (2023). "Real-Time Machine Learning Application for Formation Tops and Lithology Prediction." Offshore Technology Conference D011S005R001.
31. Michael, W., Dunham., Alison, Malcolm., J., Kim, Welford. "Predictive lithology mapping using semisupervised learning: Practical insights using a case study from New South Wales, Australia." Geophysics, undefined (2023). doi: 10.1190/geo2022-0476.1.
32. Chawshin, Kurdistan, Carl Fredrik Berg, Damiano Varagnolo, and Olivier Lopez. (2021). "Lithology Classification of Whole Core CT Scans Using Convolutional Neural Networks." SN Applied Sciences **3**(6).