GEOSCIENCES

Machine learning-based classification of petrofacies in fine laminated limestones

GALLILEU GENESIS, IGOR F. GOMES, JOSÉ ANTONIO BARBOSA, ARALY FABIANA L. DE ARAÚJO & GERMANO MÁRIO S. RAMOS

Abstract: Characterization and development of hydrocarbon reservoirs depends on the classification of lithological patterns from well log data. In thin reservoir units, limited vertical data impedes the efficient classification of lithologies. We present a test case of petrofacies classification using machine learning models in a thin interval of finely laminated limestones using pseudo-well data created over outcrops (radiometric and unconfined compressive strength logs). We tested Gaussian naïve Bayes (GNB) and support vector machine (SVM) techniques to classify eight petrofacies types, divided into two groups. The objective was to observe the capacity of some well-known models to classify petrofacies with a high-frequency vertical variation of diagenetic heterogeneities in an extreme scenario within a thin sedimentary interval. The GNB was less effective (F_1 score of 0.29), and the SVM achieved the best results in classifying the main facies patterns (F_1 = 0.47). However, the GNB performed better when the analysis was focused on distinguishing the two main groups of petrofacies. The results demonstrate that high-frequency facies variations present a challenge to the automatic identification of lithofacies, mainly due to local variations in horizontal heterogeneities (on the mm- to cm-scale) created by depositional and diagenetic processes, which impact the flow in porous media.

Key words: reservoir modeling, machine learning, petrofacies classification, thin reservoir unit, laminated limestones.

INTRODUCTION

Successful hydrocarbon extraction from conventional and unconventional reservoirs depends on constructing reliable 3D models describing various properties of the sedimentary successions (i.e. lithology, petrophysics, geomechanics, i.e) based on measurements from wells and seismic data (Bonnell & Hurich 2008, Heidsiek et al. 2020, Jones et al. 2008, Yu et al. 2008, Zhang et al. 2006). The distribution of properties in an industry-standard 3D model draws on data from wellbores (lithology succession, petrophysics, fluid types) with a spatial resolution of centimeters to meters, and seismic reflection data (2D and 3D volumes) with a spatial resolution of tens to hundreds of meters (Burchette 2012, Grana et al. 2016, Ozkan et al. 2011, Raeesi et al. 2012, Worden et al. 2018). Heterogeneities formed by depositional and diagenetic processes present a challenge for reservoir modeling, especially for carbonate rocks (Burchette 2012, Worden et al. 2018). Populating a large 3D model using the 1D measurements from wellbores (cores and logging) is usually treated using geostatistical tools (Anna et al. 2009, Correia et al. 2016, Haese 2019, Ringrose et al. 2008). Reservoir modeling is improved by information from natural exposures (reservoir analogs), which helps in understanding the patterns formed by depositional processes and heterogeneities created by diagenesis (Adams et al. 2011, Enge

et al. 2007, Questiaux et al. 2009). Reservoir analogs can provide large amounts of data, including geometries, mechanical and petrophysical properties, and fracture system properties (Anna et al. 2009, Bayer et al. 2015, Belayneh et al. 2006, Enge et al. 2007, Falivene et al. 2006, Heidsiek et al. 2020, Howell et al. 2014, Jones et al. 2008, Milad & Slatt 2019, Yan et al. 2020).

Static reservoir modeling depends on the power to predict intrinsic properties in regions far from the wellbore (Grana et al. 2016, Zhao et al. 2014), and well observations are not always sufficient to inform the predictions desired (Burchette 2012, Laubach et al. 2019), such as in thin reservoir units with great horizontal extensions. Furthermore, the manual labeling of properties of interest, either lithofacies (groups of rocks that share similar lithologic or physical characteristics), or petrofacies (groups of rocks that share similar petrographic or mineralogical characteristics) regarding variations found from site to site over the borehole data can also be ambiguous, expensive, and time-consuming (Edwards et al. 2017, Lineman et al. 1987). This challenge has led to the development of automated identification and classification computation tools that process well log observations (Hall 2016, Halotel et al. 2019, Merembayev et al. 2021, Silva et al. 2020) and core images (Chawshin et al. 2021, Lima et al. 2019, Thomas et al. 2011). These approaches allow fast and reliable identification/correlation of geological properties from well logs (Wu et al. 2018), to build sophisticated models (Ertekin & Sun 2019, Othman et al. 2021). Most of these new techniques are based on machine learning (ML) algorithms.

When modeling complex, thin reservoir units from noisy and limited data from few wells and limited well log records, more sophisticated classification algorithms such as neural networks are not applicable since they require large amounts of data for their training. In contrast, less complex algorithms are more robust when dealing with noisy or constrained training data. Even with fewer parameters or assumptions about the data, they tend to be less prone to overfitting, than more complex models. Thus, we chose GNB and SVM models, among the less sophisticated ML techniques, for the classification task (Bisho 2006, Murphy 2012). Another aspect we considered is the opportunity of comparing two algorithms with different approaches, one probabilistic (GNB) and the other discriminative (SVM), for the limited-data case studied.

Both of these models (SVM mostly) are well-established ML approaches to lithofacies classification (for SVM, see Al-Anazi & Gates 2010, Alexsandro et al. 2017, Xie et al. 2018, Deng et al. 2019, Sarkar & Majumdar 2020, Fadokun et al. 2020, Liu et al. 2020, Verma et al. 2021, Kumar et al. 2022, Gonzalez et al. 2023, for GNB, see Li & Anderson-Sprecher 2006, Horrocks et al. 2015, Babasafari et al. 2022, Nwaila et al. 2022, Nguyen et al. 2022). However, the literature presents very few studies with applications of these methods in the task of classifying petrofacies (López & Thomas 2009, Duarte et. al 2023, Silva et al 2020). The situation becomes particularly challenging when searching for research papers that apply SVM and/or GNB for petrofacies classification in data-limited settings, as no such studies were found in the literature.

This work studied the capacity of ML techniques to treat a specific scenario comprising a thin unit of laminated limestones with a high-frequency vertical variation of petrofacies resulting from depositional and diagenetic processes. It represents a critical case study involving a relatively thin and continuous interval of carbonate rocks used to verify the challenge faced in the petrofacies distribution due to the horizontal extension of the sampled interval. We tested the performance of GNB and SVM in the automated identification of petrofacies in a thin limestone interval. The main

goal was to treat the petrofacies classification problem in an extreme scenario of scarce data, a condition in which these algorithms have not yet been tested. The choice of these two algorithms also aimed to compare the two completely different approaches, one probabilistic (GNB) and the other discriminative (SVM), which can provide insights into the performance of different modeling techniques in the context of the specific reservoir units and data constraints.

The succession studied is the C6 unit, a stratigraphic interval of laminated limestones in the upper part of the Crato Formation of the Araripe Basin in northeastern Brazil (Fig. 1a-c). We used the petrofacies classification in previous works (Araujo et al. 2020, Ramos et al. 2020), which integrated
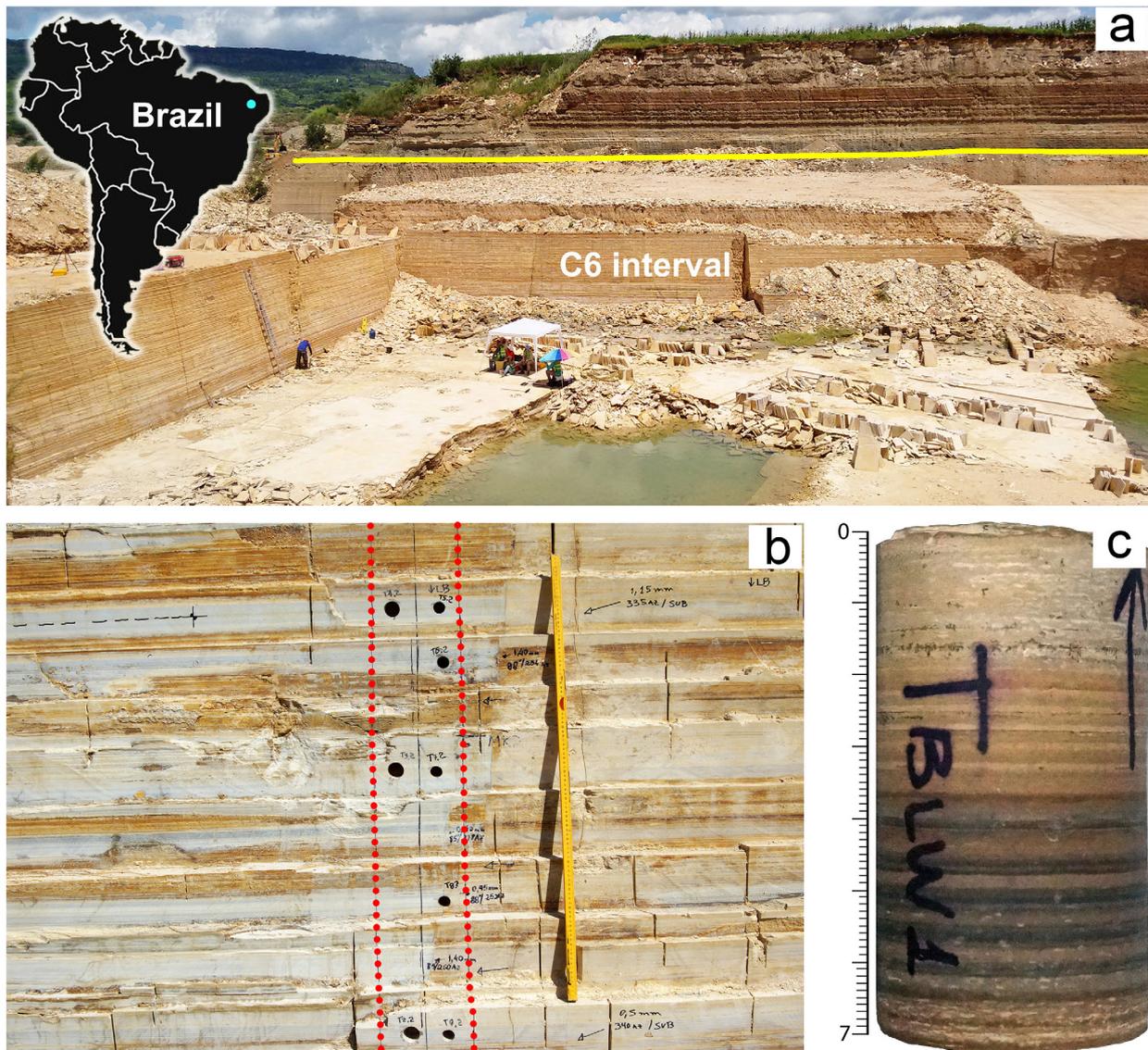

Figure 1. C6 Interval of laminated limestone from the Crato Formation, Araripe Basin. a) View of the interval of limestones studied in a quarry, Nova Olinda region. b) A vertical strip (dotted red lines) used to define the acquisition of data and samples emulating a vertical well (pseudo-well) (yellow stick = 1 m). c) A plug showing the fine laminations that were classified in thirteen petrofacies. The greyish and yellowish colors indicate variations caused by depositional and diagenetic processes (Araujo et al. 2020).

lithological, depositional, and diagenetic characteristics. Petrofacies complement the lithofacies concept, which can integrate properties like porosity and permeability (Bhattacharya et al. 2005, Cao et al. 2020, 2021, Jardim et al. 2011, Kadkhodaie & Kadkhodaie-Ilkhchi 2018). This succession was chosen because it exemplifies two assessment challenges in reservoir modeling: a thin vertical interval and the high-frequency vertical variation of petrophysical properties of the fine laminations (mm-scale). This type of high-frequency variation in depositional and diagenetic properties strongly influences reservoir quality (over the mm- to cm-scale) (Mikes et al. 2006, Creusen et al. 2007, Likuan et al. 2021), and its integration through up-scaling techniques for reservoir characterization represents a major challenge (Mikes et al. 2006, Heidsiek et al. 2020). The study tests the performance of the ML models on the extreme case, with limited data, and demonstrates that the vertical variation of the petrofacies is hard to resolve, even in a unit with apparent good horizontal continuity of lithofacies.

Geological Setting and Pseudo-Wells Dataset

The geological data used in this study were collected from outcrops in two quarries in Nova Olinda, Araripe Basin, Ceará State, Brazil (Fig. 1). Three pseudo-wells were built over vertical exposures with lateral continuity of tens of meters (Araujo et al. 2020). The C6 interval studied is a succession consisting of finely laminated limestones, composed predominantly of micritic calcite (mudstones), with local occurrences of dolomite and silica (Araujo et al. 2020, Miranda et al. 2018). The C6 interval presents a regional distribution and is part of the first post-rift sequence of the basin (Assine et al. 2014, Neumann et al. 2003). These rocks were previously studied as an analog for tight fractured reservoirs (Miranda et al. 2018, Santos et al. 2015) and as an analog for lithofacies found in the Brazilian pre-salt fields (Barra Velha Formation) (Catto et al. 2016). The C6 interval contains centimetric- to metric-scale vertical calcite veins, shear fractures and centimetric vertical stylolites. Furthermore, metric-scale joints are related to late exhumation (Miranda et al. 2018). The laminations are 3–5 mm thick, and the variation in depositional conditions (climate, salinity, and sedimentation rates) resulted in a varying content of organic matter (OM) in the laminae (Fig. 1). Variation in the sedimentation rate also influenced matrix cementation during the early stages of diagenesis (Araujo et al. 2020, Heimhofer et al. 2010, Osés et al. 2017), resulting in cemented sets of laminations with early silica formation. Early silicified sets of laminae also contain higher contents of OM and pyrite. These levels present a grey-greenish color (G2 group of laminations). The C6 succession also presents lamination sets with iron oxides, and a minor content of silica and OM, with a yellowish color (G1 group of laminations). The variation in porosity and mechanical strength between the two groups of laminations is expressive. The G1 group presents more dissolution features, higher porosity, and fewer deformational structures (Araujo et al. 2020).

    The study used data from three vertical pseudo-wells created with high-resolution stratigraphic descriptions of the petrofacies from two quarries located one kilometer apart (Figs. 2a,b and 3) Araujo et al. (2020). The first and second pseudo-wells located in panel 1 (PW1 and PW2 – Idemar Quarry) are separated by 9.95 m, and the third pseudo-well was located in panel 2 (PW3 – William Quarry). Petrofacies can be defined by integrating the rock sample observations and well logs (Cao et al. 2021, Jardim et al. 2011). This approach considers the characterization of lithofacies (lithology, sin-depositional structures) and other physical aspects like diagenetic features (cementing and
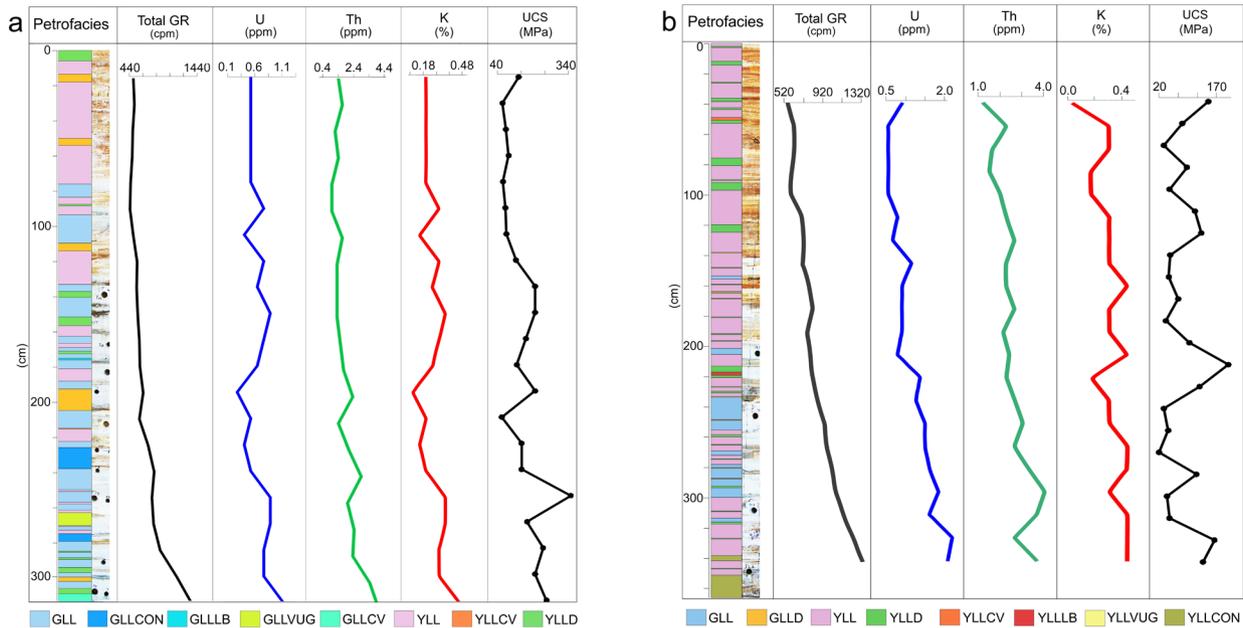
Figure 2. Composite well logs with petrofacies, Total GR, U, Th, K and UCS data of the two pseudo-wells acquired in panel 1 (a) PW1 and b) PW2 – Idemar Quarry). Spacing for the acquisition of geophysical data was 15 cm.
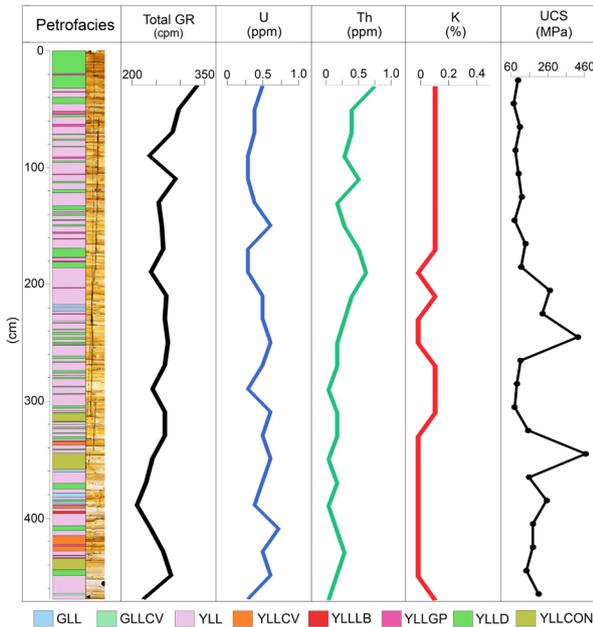

Figure 3. Composite well logs with petrofacies, Total GR, U, Th, K and UCS data of the pseudo-well acquired in panel 2 (PW3 – William Quarry). Spacing for the acquisition of geophysical data was 20 cm.

concretions), petrophysics (porosity and permeability), and mechanical properties (Araujo et al. 2020, Cao et al. 2021, Watney et al. 1998, Gómez 2020, Jardim et al. 2011). The main advantage of the petrofacies approach for sedimentary succession classification is that it can constrain the distribution of parameters directly related to fluid flow. The petrofacies classification provided by Araujo et al. (2020) for the C6 interval integrates sedimentological, diagenetic, and sin-depositional features. The authors defined thirteen (13) petrofacies in two main groups, seven in the beige-yellowish G1 group and five in the grayish G2 group, as shown in Table I.

Table I. Description of the two petrofacies groups identified in the laminated limestone interval, with the main characteristics of depositional and diagenetic features (Araujo et al. 2020).

Petrofacies Group	Depositional and Diagenetic Characteristics	Petrofacies code
G1	Yellow Laminated Limestone	YLL
	Yellow Laminated Limestone with Loop beddings	YLLLB
	Yellow Laminated Limestone with Convolution	YLLCV
	Yellow Laminated Limestone with Dissolution	YLLD
	Yellow Laminated Limestone with Vug	YLLVUG
	Yellow Laminated Limestone with Concretion	YLLCON
	Yellow Laminated Limestone with Gypsum Vein	YLLGP
G2	Gray Laminated Limestone	GLL
	Gray Laminated Limestone with Loop Beddings	GLLLB
	Gray Laminated Limestone with Convolution	GLLCV
	Gray Laminated Limestone with Dissolution	GLLD
	Gray Laminated Limestone with Vugs	GLLVUG
	Gray Laminated Limestone with Concretion	GLLCON

The succession shows a high-frequency alternation between the G1 and G2 petrofacies components. The G2 petrofacies dominate the basal part of the succession. Stratigraphic information was acquired with a vertical limit of 5 mm for the definition of the individual petrofacies, resulting in a high resolution stratigraphic description in the pseudo-wells (Araujo et al. 2020). The pseudo-wells data also included: 1) - gamma ray logs acquired in situ using a portable gamma spectrometer. Measurements comprise the total gamma emissions and the potassium (K), thorium (Th), and uranium (U) contents (Araujo et al. 2020), and 2) - in situ unconfined compressive strength (UCS) measurements acquired with a portable sclerometer (Schmidt Hammer N type). The sampling spacing for the gamma and UCS tests was 15 cm for the two pseudo-wells in panel 1 and 20 cm for the pseudo-well in panel 2 (Araujo et al. 2020). Figs. 2 and 3 show the composite well data with the gamma logs, which are related to the mineral composition of the strata, and the UCS data, which are related to its mechanical properties. Previous tests indicate that these logs could help to distinguish between the petrofacies because the gamma measurements are sensitive to variations in diagenetic features (primary and authigenic mineral content), and the UCS shows good relation to properties like density and cementation index.

Because of the high resolution used to describe the vertical profiles (mm-scale), the sampling of gamma and mechanical strength (15 to 20 cm spacings) resulted in these measurements being captured and integrated for only eight petrofacies. That result demonstrated the first challenges of capturing information for this type of succession (Araujo et al. 2020). Thus, the research considered the eight petrofacies which were integrated with the physical logs for ML-based processing: GLLCV, YLL, GLLVUG, GLL, YLLCON, YLLLB, YLLCV, and YLLGP. Fig. 4a-c shows the distribution of each petrofacies in the total vertical succession sampled. A substantial imbalance is observed in the dataset, as YLL (52.31%) and GLL (23.08%) are the most frequent facies.
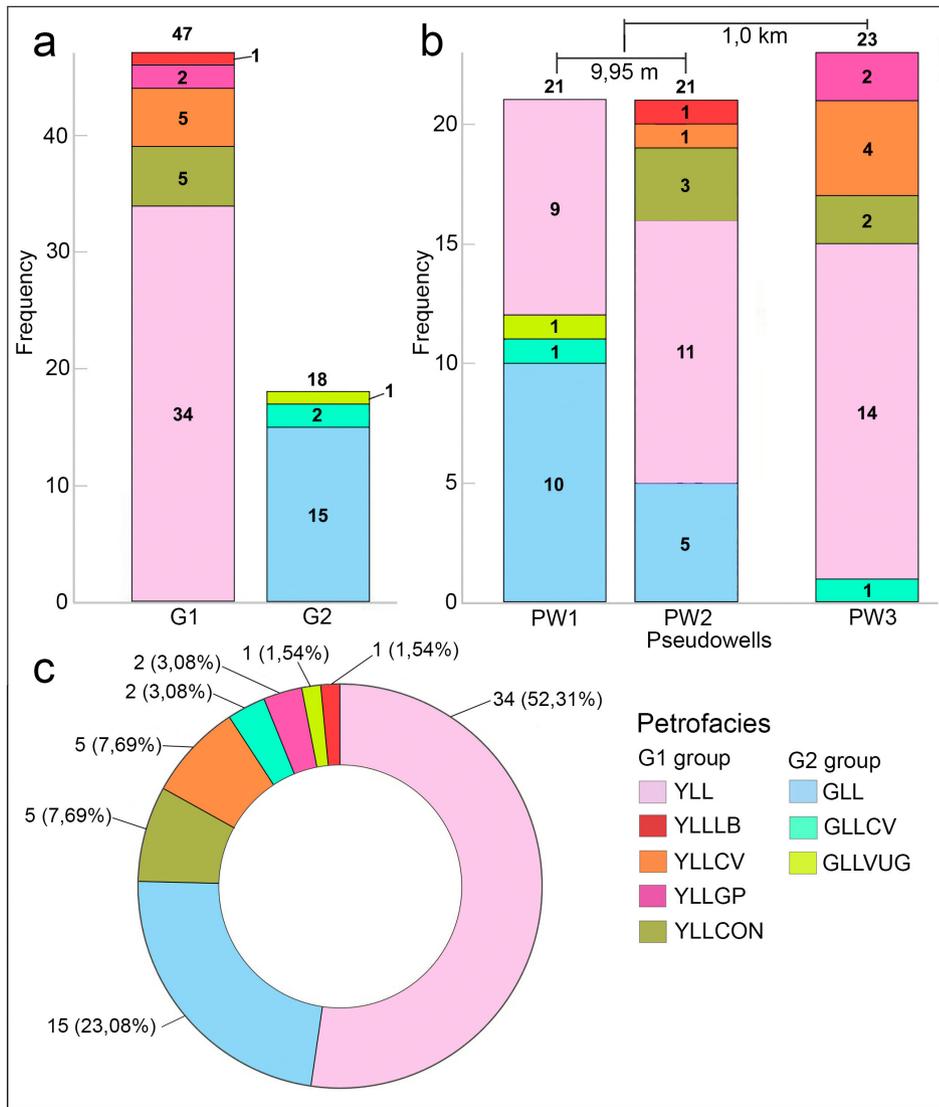
Figure 4. a) Distribution of the eight petrofacies from both groups G1 (grey) and G2 (yellow) tested in the vertical space of the pseudo-wells. b) Distribution of petrofacies in the vertical profiles of the three pseudo-wells. c) Distribution of petrofacies classes in the dataset.

MATERIALS AND METHODS

Gaussian Naïve Bayes

NB is one of the most traditional ML algorithms for classification purposes. The method is based on the Bayes' Theorem and assumes that features are independent (Witten & Frank 2002).

The prediction of class $\hat{y}$ with the NB algorithm is given by the maximum *a posteriori* probability (MAP) estimate:

$$\hat{y} = argmax_{(y_k)}\left[ln[P(y_k)] + \sum_{i=1}^{m}ln\left(P(X_i|y_k)\right)\right], \quad with \quad k = 1,2,...,K \qquad (1)$$

Assuming that the features **X** follow a normal distribution, a particular case of NB—Gaussian naïve Bayes (GNB)—is used, in which the conditional probability $P(X_i|y_k)$ is given by:

$$P(X_i = x|y_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}}e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}} \qquad (2)$$

Support Vector Machine

SVM is a fundamentally discriminative classification model whose idealization is based on statistical learning frameworks (Cortes & Vapnik 1995, Vapnik & Chervonenkis 1971, Vapnik 1995).

For a binary classification problem with a linearly separable dataset, the SVM looks for the hyperplane (Eq. 3) that separates the classes with the maximum distance or margin (Eq. 4) with respect to a subset of training points called support vectors.

$$W^T X + b = 0 \tag{3}$$

$$\rho = \frac{2}{||W||} \tag{4}$$

The optimal hyperplane maximizes $||W||$, which can be found by solving the following equation:

$$min_{(W,b)} \frac{1}{2} W^T W + C \sum_{i=1}^{m} \xi_i \quad s.t. \quad y_i \left(W^T X_i + b \right) \geq 1 - \xi_i, \quad \xi_i \geq 0 \tag{5}$$

In this formula, the parameter C controls the magnitude of the margin and, consequently, the precision of the model. Furthermore, ξ indicates the number of support vectors that the model must use to construct the regression function. Figure 5 shows a cross-plot of the features obtained through the pseudo-wells. Most features indicate a reasonable separation between the wells. However, the separation of the features of PW3 from PW1 and PW2 is notable, indicating a spatial dependence of the measured features and, therefore, of the lateral variation of geological properties. This will impact upon the performance of the models depending on the training and testing configurations. Besides the features mentioned before—total gamma ray counts, separated K, U and Th counts, and UCS measurements—we also use the depth of each measurement. Normally, layered rocks exhibit good lateral continuity of some parameters, and this was also observed in the C6 interval. (Araujo et al. 2020) found three distinct mechanical zones in the C6 interval, which information was used to improve the occurrence probability of the petrofacies through the pseudo-well sections.

ML Training and Evaluation

We applied the algorithms to three different classification sets, as shown in Fig. 6. In each classification set, two pseudo-wells were used for training and the third was used for testing. k-fold cross-validation was used to evaluate the model's performance on the training data and then the model was applied to the test data. The SVM hyperparameters were tuned using sklearn GridSearchCV. The GNB represents a nonparametric model and, therefore, didn't need a hyperparameter tuning scheme. We determined the model's generalization capacity by comparing the accuracy, precision, recall, and F_1-scores it obtained over the specified training data, and then the test data.

Besides the asymmetry of the petrofacies distribution in the three pseudo-wells, another important aspect is the absence of some petrofacies types in the training dataset that appear in the test dataset, as shown in Fig. 7. This occurs due the complex vertical and horizontal distribution of the petrofacies, and demonstrates real conditions found in sedimentary successions. For example, petrofacies GLLVUG appears only in PW1, YLLLB which is found only in PW2, and YLLGP is present only in PW3.

RESULTS

This section describes the results of training data cross-validation using all the features of the dataset, and the combination of features that yielded the best classification performance. We also present the results for the three classification scenarios involving the three wells (Fig. 6) using the GNB and SVM models.
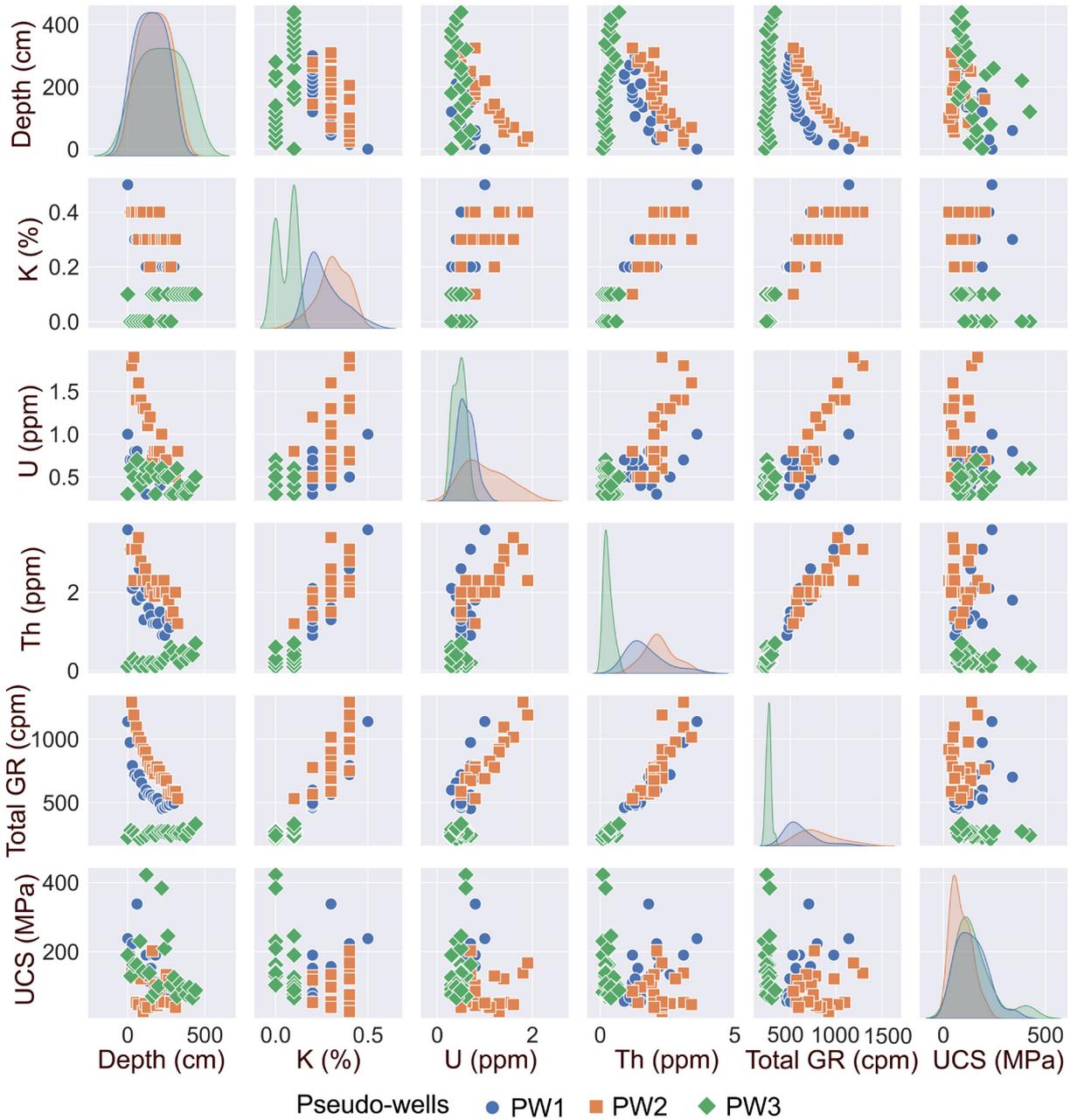

Figure 5. A cross-plot showing the relationships between all features (Depth, K, U, Th, Total GR and UCS) collected from the three pseudo-wells PW1, PW2 and PW3.
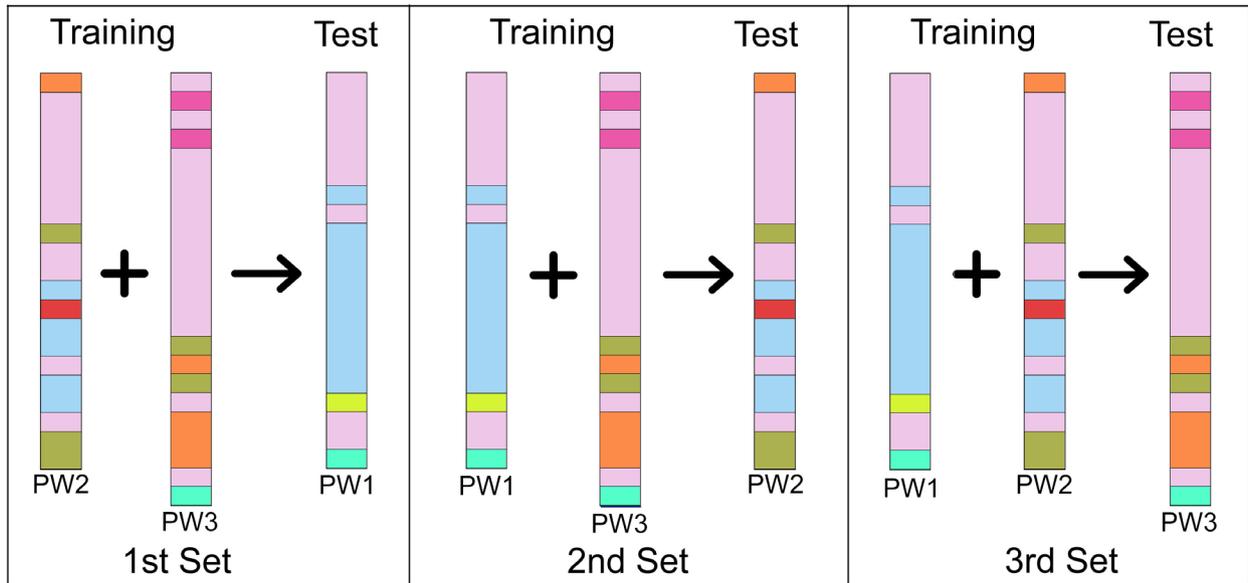
Figure 6. 1st set: PW2 and PW3 used for training and PW1 used for model evaluation. 2nd set: PW1 and PW3 used for training and PW2 used for model evaluation. 3rd set: PW1 and PW2 used for training and PW3 used for model evaluation (See the color code for petrofacies in Figures 2 and 3).
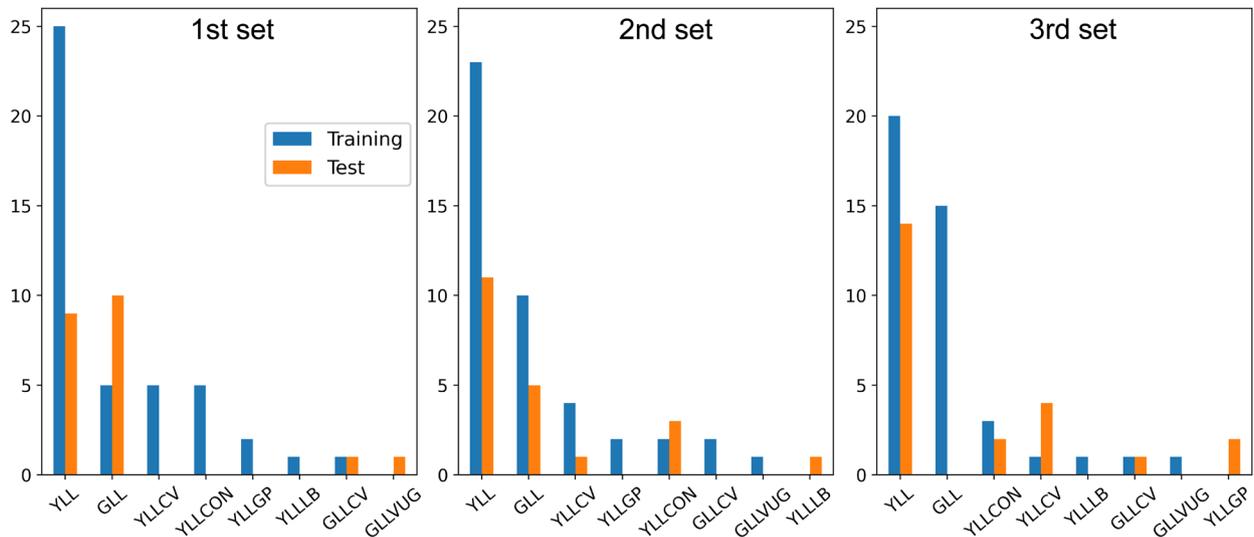

Figure 7. Petrofacies distribution in the training and test datasets for all classification sets (See Petrofacies codes in Table I).

GNB Performance

We conducted an ANOVA to compare different combinations of features and the performance of the GNB model under 5-fold cross-validation. The results of this analysis are presented in Tables II and III.

The cross-validation results for the GNB model based on the Depth, K, and U features are shown in Fig. 8.

Figure 9 shows the means and standard deviations obtained by 5-fold cross-validation with the three training sets for both cases: using all features, and using the best configuration of features.

Table II. ANOVA performed on the features (Depth, K, U, Th, Total GR and UCS), sorted by decreasing p values.

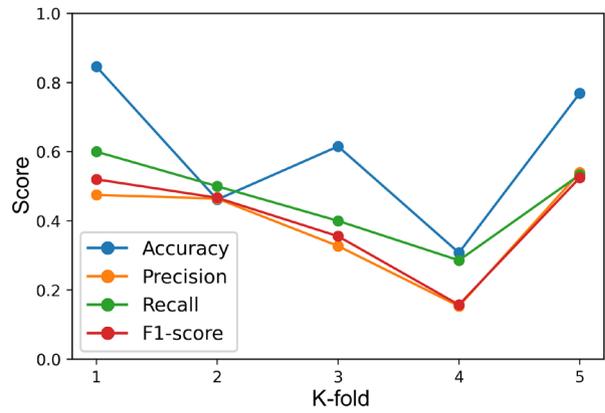| Feature	F-value	p-value
Depth	7.027236	0.000012
K	3.833263	0.002801
U	3.035999	0.012030
Th	3.011069	0.012595
Total GR	2.949709	0.014104
UCS	1.291329	0.275910

Figure 8. 5-fold cross-validation for the GNB model based on the Depth, K, and U features.

Table III. Feature combinations and their respective mean accuracy, precision, recall, and F_1 score in the cross-validation of GNB classification.

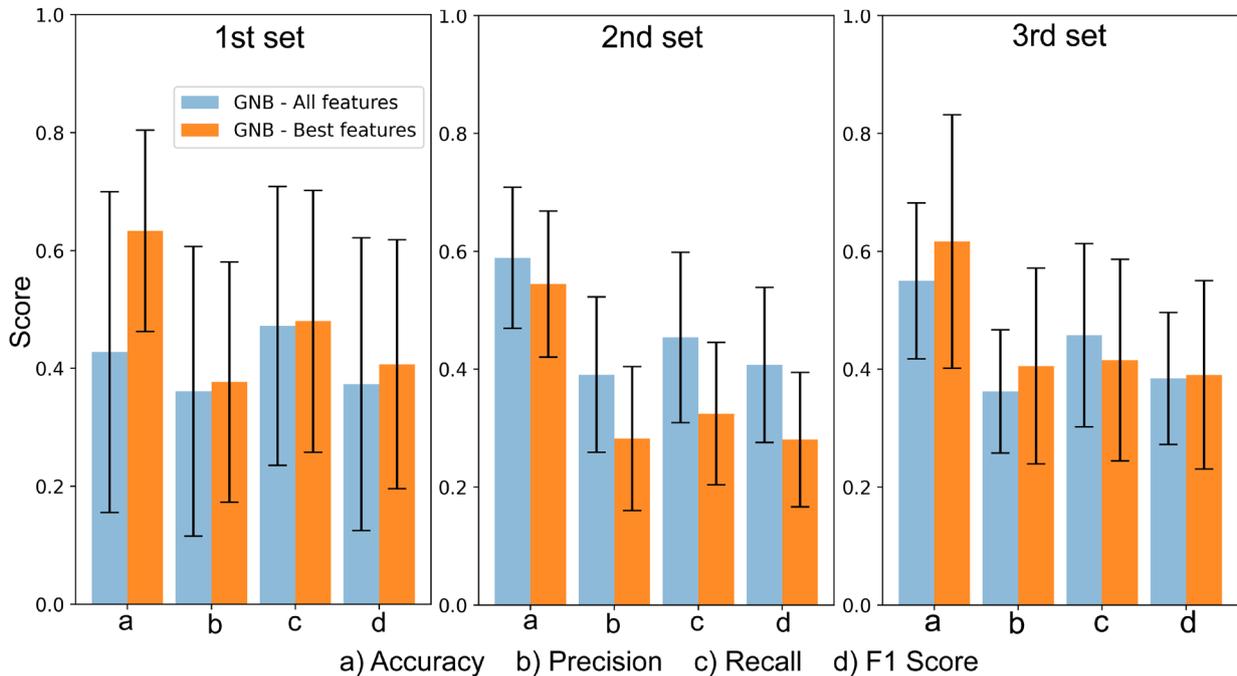| Feature Blend	Accuracy	Precision	Recall	F1 Score
Depth, K, U	60.00	39.20	46.38	40.48
Depth, K, Total GR	58.46	36.78	47.11	38.02
Depth, K	60.00	38.37	43.87	37.71
Depth, K, U, Total GR	55.38	35.23	44.67	36.96
Depth, K, Th	56.92	33.06	43.78	34.35

Figure 9. Means and standard deviations of the cross-validation (k-fold = 5) for the GNB models with all features and best features in the training sets.

Figure 10 shows the classifications obtained by each GNB model considering the test data, according to the configuration described in Fig. 7.

Figure 11 shows the confusion matrices used to analyze the performance of the three model configurations, for petrofacies classification.

SVM Performance

We processed the SVM models using the same procedures for feature selection as used for the GNB models, and the five best mean results are shown in Table IV, sorted by F_1 score.

We used the results in Table IV to choose the best SVM model: the set processed with all features. The cross-validation results for this model are shown in Fig. 12.

The performance metrics for the SVM and best GNB processing on the test data across all training datasets are shown in Figure 13, and Figure 14 shows the confusion matrices for the SVM and best GNB models.

The facies classification results obtained with the models are shown in Figure 15, along with the actual petrofacies distribution: a) the original vertical petrofacies succession interpreted in the field; b) the "actual" classification used by this research, which presents a coarser vertical classification of the petrofacies succession due to the sampling spacing of the features; and c and d) the classification produced by the GNB and SVM models. The overall predominance of the petrofacies in the three wells reveals that the C6 Interval in the PW3 site is dominated by petrofacies of the G1 group, which presented a problem for the classification. The models were evaluated on their capacity to achieve classification against the actual distribution of the eight petrofacies in the two groups (G1 and G2), as well to classify the strata for the two groups in general (Fig. 15).

Table V shows the proportion of the petrofacies classified regarding their relationship to the two groups against the actual proportion present in the three wells, according to the results obtained by the best GNB and SVM models.

Table IV. Feature combinations and respective means of accuracy, precision, recall, and F_1 score, from SVM processing.

Feature Blend	Accuracy	Precision	Recall	F1 Score
(Depth, K, U, Th, Total GR, UCS)	60.00	37.12	38.48	34.85
(Depth, K, U)	63.08	34.38	38.56	34.60
(Depth, K, Th, Total GR)	61.54	32.95	38.21	33.81
(Depth, K, U, Th)	60.00	33.56	36.87	32.74
(Depth, K, Th, Total GR, UCS)	60.00	34.97	35.29	32.70

Table V. Comparison of proportions of the classified petrofacies within groups G1 and G2 provided by the models for the three pseudo-wells against the actual proportions of each group.

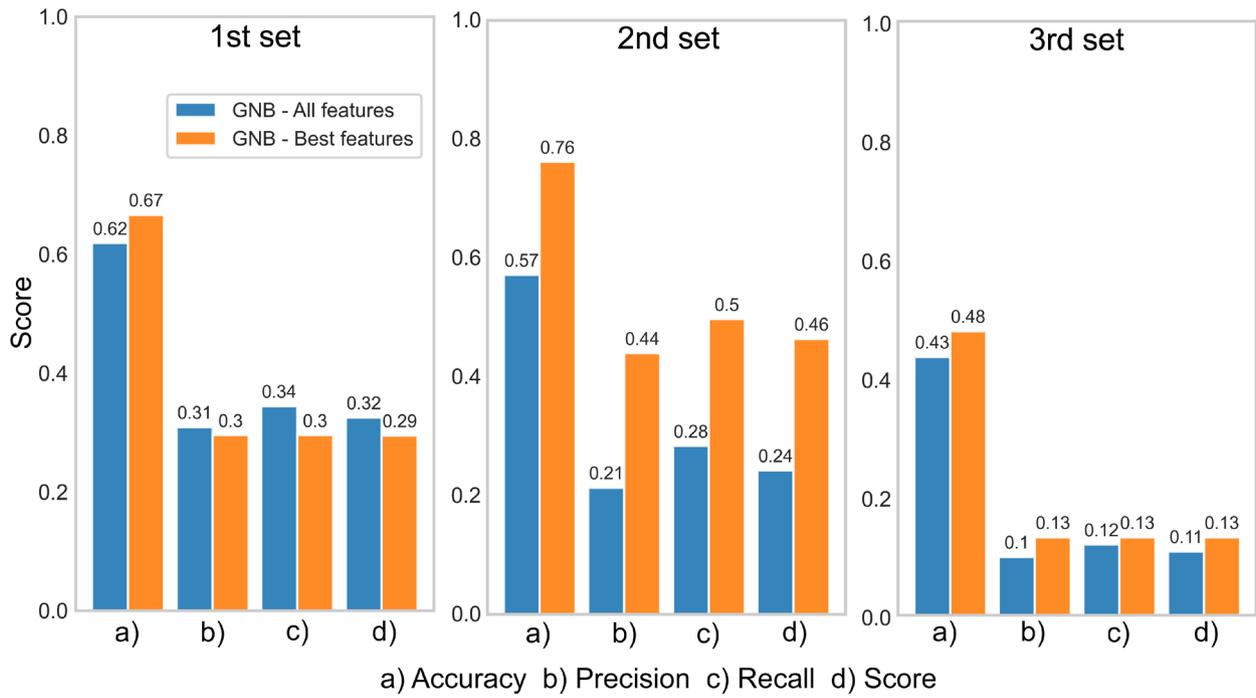| Group	1st set			2nd set			3rd set		
	Actual	GNB	SVM	Actual	GNB	SVM	Actual	GNB	SVM
G1 (Yellow)	43%	48%	38%	76%	67%	62%	96%	61%	52%
G2 (Gray)	57%	52%	62%	24%	33%	38%	4%	39%	48%

Figure 10. Results of the GNB models with the three test sets, for all features and for the best configuration of features.
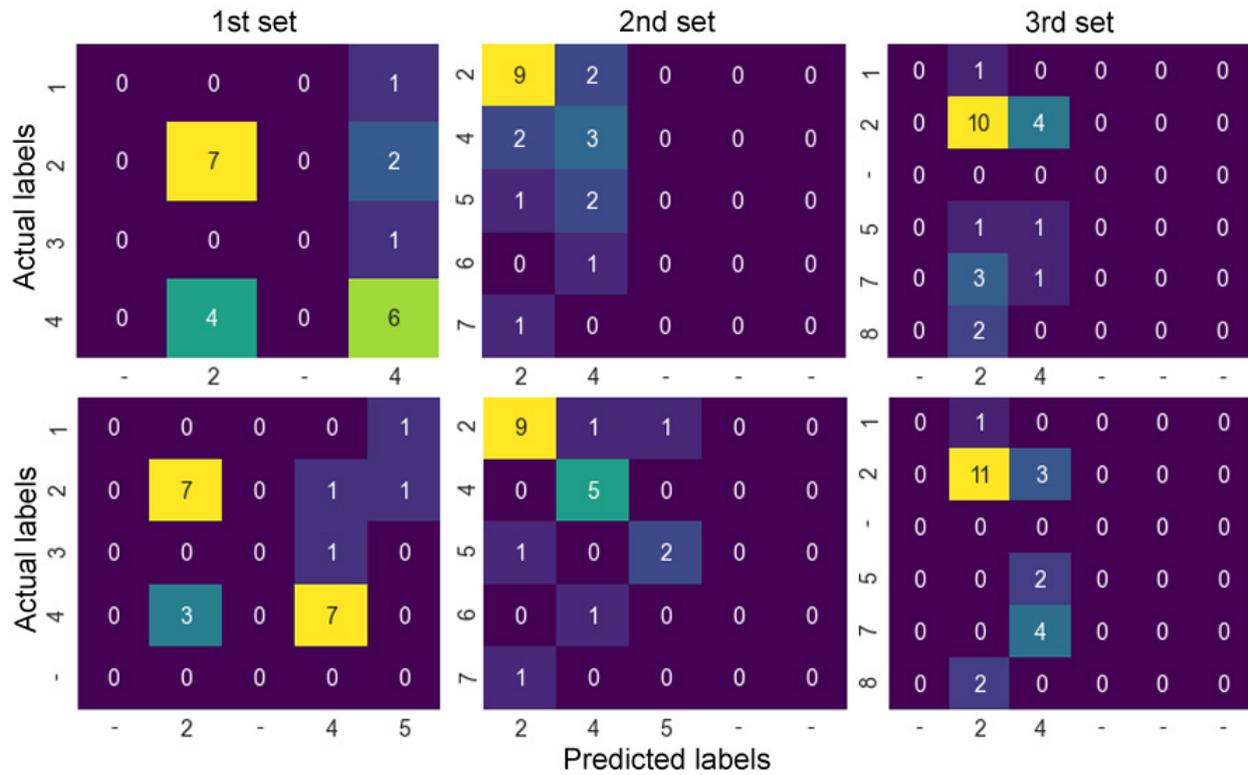

Figure 11. Confusion matrices of the GNB models for the three classification sets. Top: model that used all features. Bottom: model that used the best features. Petrofacies dictionary: 1-GLLCV, 2-YLL, 3-GLLVUG, 4-GLL, 5-YLLCON, 6-YLLLB, 7-YLLCV, and 8-YLLGP.
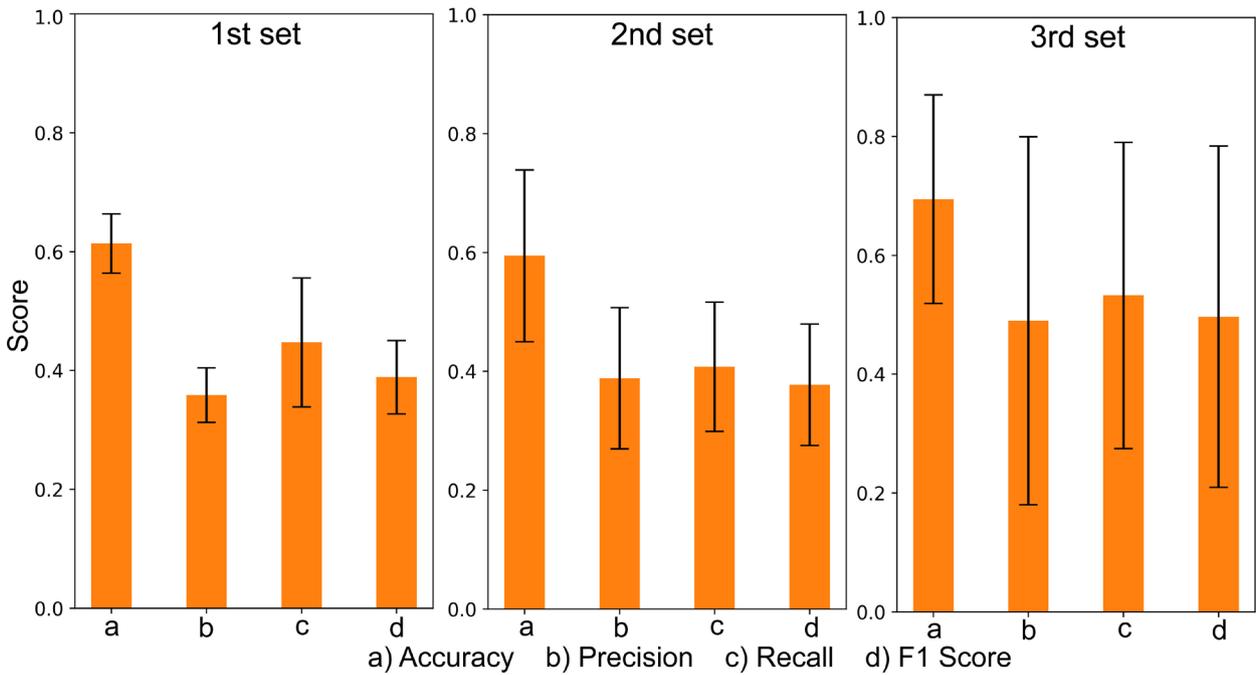
Figure 12. Means and standard deviations of the 5-fold cross-validation obtained by SVM processing using all the features.
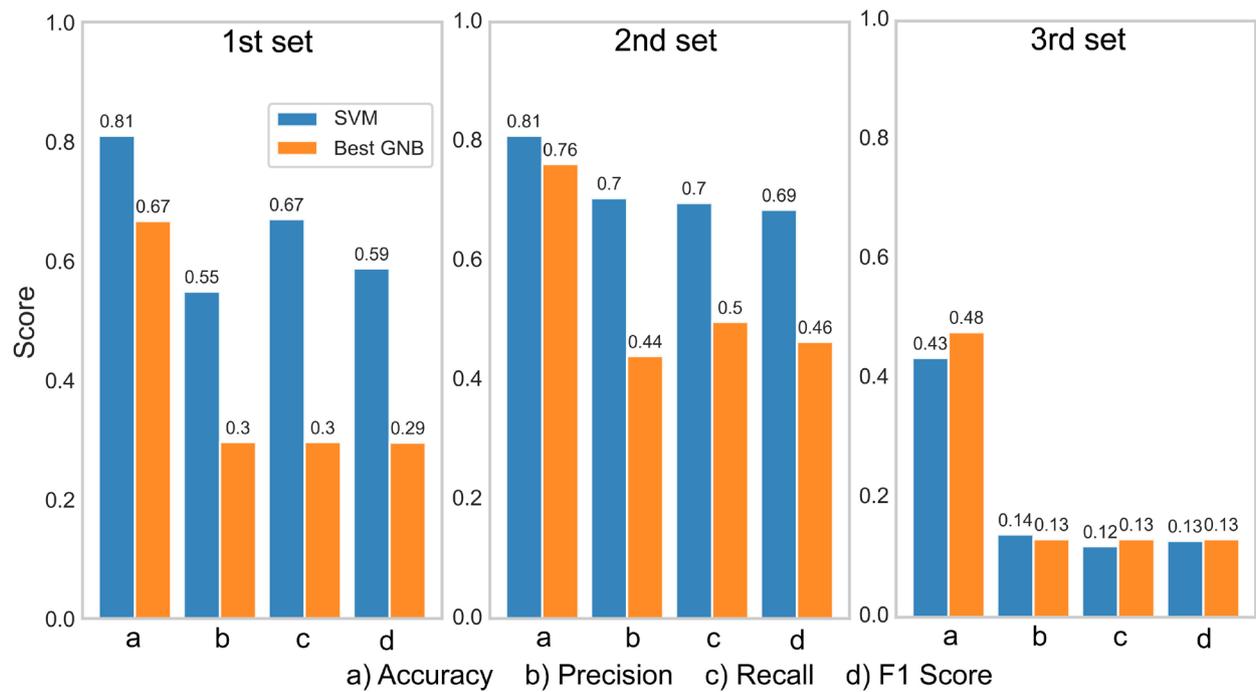

Figure 13. Performance metrics for the SVM and best GNB models processed with the three test sets.
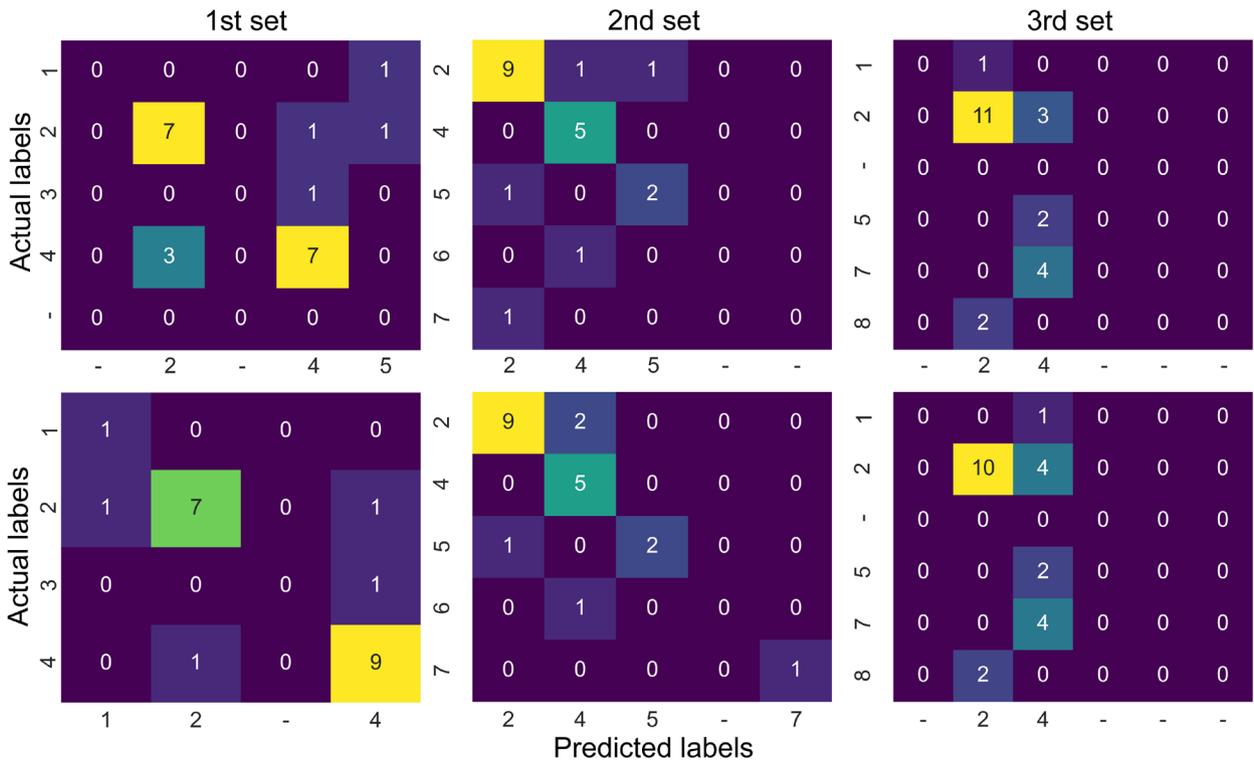
Figure 14. Confusion matrices for the best GNB model (top) and the SVM model (bottom). Petrofacies dictionary: 1-GLLCV, 2-YLL, 3-GLLVUG, 4-GLL, 5-YLLCON, 6-YLLLB, 7-YLLCV, and 8-YLLGP.

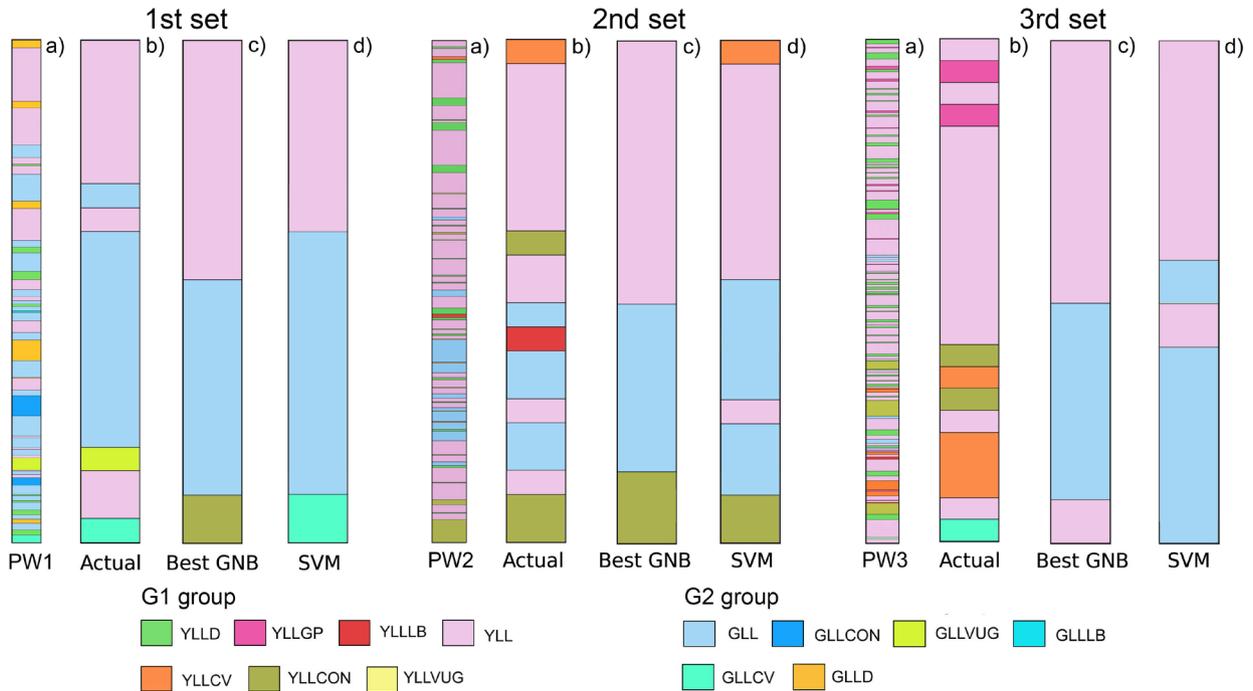
Figure 15. Description of the pseudo-wells and the classifications by the models. a) Original distribution of the thirteen petrofacies in the vertical profiles of the pseudo-wells. b) Actual distribution of the eight petrofacies integrated with the physical parameters used for classification. c) GNB results, and d) SVM results. See the color code for all the petrofacies in Figures 2 and 3.

DISCUSSION

Figure 9 shows that the GNB model using the best features (best GNB) performed better on the first and third training dataset, which demonstrates the importance of feature selection in this case. For both models, the cross-validation results are similar across all training datasets. The positive impact of using the best features on the performance of the GNB model can be observed in the results obtained from the test data (Fig. 10). The classification accuracy on the first set increased by 8% while there was no improvement in the other metrics. In the second set, the accuracy improved by just over 33% in the best GNB model, while the precision, recall, and F_1 score more than doubled. The third classification set showed slight improvements in all metrics.

Furthermore, the best GNB model performed better on the test data than on the training data on the first and second classification sets, and lower performance on the third set. The training data used for the cross-validation between the first and second sets, PW1 and PW2, were from locations 9.95 m apart, and therefore there was better lateral continuity of geological characteristics between them. In the two other configurations, the data used for PW3 was 1.0 km distant from the data from PW1 and PW2 (Fig. 6), and the results showed that the lateral variation in the geological characteristics within the C6 interval made petrofacies identification more difficult, as expected. This also explains the high standard deviations obtained in the cross-validation of the training set. Moreover, the limited amount of data available also impacted the performance of the models with the training sets. The small data volume explains the improvements in the metrics, particularly accuracy, obtained from test data versus those obtained from cross-validation.

Analyzing the model performance showed that, in the first set, which was devised to determine the petrofacies distribution in PW1 based on data from PW2 and PW3, the training set was unbalanced because the petrofacies YLL has a frequency greater than the sum of all others. The test set comprised only four petrofacies, with the following frequencies: YLL (9), GLL (10), GLLCV (1), and GLLVUG (1) (Fig. 7). The frequency of petrofacies GLL in the test data was twice that of the training data. However, as it has different geological characteristics from the G1 Group, the model was able to differentiate them, despite the differences in sampling and representation. With these characteristics, the model training tends to generate a bias for the YLL and GLL petrofacies, which can be seen in the confusion matrix in Fig. 11. Regarding the petrofacies GLLCV and GLLVUG, the latter occurs only in the test data, making its identification impossible, while the former occurs exactly once in both sets, which makes its identification probability very low. However, due to their limited occurrence in the test data, the failure in the classification of these petrofacies did not have a great impact on the evaluation metrics, especially on the accuracy (Figs. 7 and 10).

In the second test set for classification in PW2, the spatial distribution of the pseudo-wells had the same characteristics used in the first configuration (Fig. 6). The arrangement of classes in the training and testing data was also very similar to the first configuration (Fig. 7). The frequency of the YLL class in both datasets is much higher than that of the other classes, which creates the same bias previously described. However, the GLL and YLLCV petrofacies were more frequent in the training data than in the test data. This tended to improve the model learning for these classes, thereby improving their performance in the subsequent classification task. These results explain why the model had better performance with the second training set, and why the improvement in the metric scores due to the use of the best features was more significant in this case (Fig. 10).

The third testing scenario proved more challenging: the training pseudo-wells (PW1 and PW2) show more lateral correlation due to their proximity, and they show significant geological lateral differences from the target pseudo-well (PW3), as expected, due to the distance between them. The training dataset (Fig. 7) was dominated by the YLL (20) and GLL (15) petrofacies. The other classes each occurred once, except for YLLCON petrofacies, which occurred three times. The test set was dominated by petrofacies YLL (14), while petrofacies GLL was absent and the other petrofacies presented a very low frequency in the PW3 set. Petrofacies YLLGP was absent from the training data. This scenario led to a higher bias effect due to the differences between the training and test sets. Thus, it is expected that the model tended to emphasize the occurrence of petrofacies YLL and GLL, which explains the overall inferior performance for petrofacies classification compared to the previous model tests. The processing showed a slight improvement in classification with the use of the best features.

Figure 11 shows that, for all classification configurations, the model that used all features was able to identify only the most frequent classes (YLL and GLL), resulting in a strongly biased classification. Using only of the best features, this bias was reduced and the model could correctly classify the YLLCON petrofacies, which illustrates the improvement in model performance for this feature selection (Fig. 11).

Table IV demonstrated that the best performance of the SVM model was obtained using all features. A comparison with Table III reveals similar results between the SVM and the best GNB. The SVM cross-validation results are also similar to those for the best GNB (Fig. 9). As was done in the cross-validation task, the amount of data used for training in each fold was indeed less than used in the sets shown in Fig. 7, and a comparison of these results reveals that using less data results in more similar performances of the SVM and GNB models. The reverse is also true, as shown by the performance of the SVM and the best GNB models through the classification sets (Fig. 13). When the amount of data available for training was larger than the amount available for cross-validation, the SVM model outperformed the GNB model running through the first two classification sets, especially regarding the precision, recall, and F_1 score metrics. Through the third training set, both models had very similar performances, with a slight advantage for the best GNB in terms of accuracy. These results are linked to the characteristics of the features used to build the models and their variations, which are controlled by the geological properties, as discussed above. As this application deals with unbalanced labels, the most suitable metric for evaluating models is the F_1 score. The SVM model obtained an average F_1 score of 0.47, while the best GNB was 0.29.

Figure 14 shows that, in general, the SVM model presented better results than the best GNB model for the first two classification sets. Especially considering the data imbalance, the model was able to correctly classify the occurrences of the two petrofacies with the lower frequencies: GLLCV (1) in the first classification set, and the YLLCV (7) in the second. Furthermore, the model showed better performance in determining the most frequent petrofacies, classifying 7 out of 9 YLL (Group 1) petrofacies and 9 out of 11 GLL (Group 2) petrofacies in the first classification set, and 9 out of 11 and 5 out of 5, respectively, in the second. The third classification set was challenging for the models due to the distance between the pseudo-wells and the consequent lateral variation of geological properties, and because occurrence and proportion of petrofacies in PW3. Thus, the greater bias observed in the GNB model may have improved its performance, which could explain its better performance in the classification of petrofacies in the third set used.

Previous classifications of C6 interval laminated limestones considered a small number of lithofacies, or microfacies. Neumann et al. (1999) divided the C6 interval into 5 microfacies: 1 – parallel and wavy-parallel laminations with loop beddings, 2 – parallel laminations with peloids, 3 – parallel and wavy-parallel laminations with micro-slumps, 4 – parallel and wavy-parallel lamination, and 5 – parallel and wavy-parallel laminations with ostracods. Catto et al. (2016) considered mineralogy, microscopic textures, and influence of microorganisms and suggested four microfacies types: 1 – planar laminated, 2 – crustiform, 3 – nodular, and 4 – rhythmic (interbedded sub-millimetric to millimetric lenses of micritic calcite, organic matter, and clay minerals). Osés et al. (2017) analyzed the preservation characteristics of microfossils in the laminated limestones of C6 interval and pointed to the influence of sedimentation rate on the alteration of organic matter as a key factor to the early diagenesis of the laminated limestones. They proposed two dominant microfacies: 1 – BL beige limestones and 2 – GL grey limestones. These authors suggested that GL differs from BL due to its pyrite, argilominerals, and OM contents. They also observed that GL microfacies present more structures like convoluted laminations, wavy laminations, and micro-faults. The classification of petrofacies adopted by the present research (Araujo et al. 2020) was built on the integration of the sedimentological aspects with chemical properties, diagenetic features and mechanical parameters (UCS) and aimed to provide information for reservoir characterization (dissolution zones, cemented zones). The differences observed between PW1/PW2 and PW3 are related to the variation in depositional conditions and diagenetic processes (local and regional effects), including the sedimentation rates and composition of sediments across the depositional system (content of clay minerals and OM). The sedimentation rate influenced the eo-diagenesis and the compaction of the deposits (Osés et al. 2017). For example, the petrofacies classification (Araujo et al. 2020) considered the occurrence of laminations bearing horizontal gypsum veins in the PW3, which were probably created by natural hydraulic fracturing linked to the proximity of fault zones—a localized diagenetic aspect (Araujo et al. 2020, Celestino et al. 2020). Based on studies of the C6 interval discussed herein, one can argue that regarding only the mineralogical composition of the succession the C6 interval can be divided into roughly two main lithofacies types. Thus, the criteria used here to improve the interpretation of reservoir characteristics was much more sophisticated and improve the characterization, but prove much more challenging for the automatic classification process. The research proved that the limited data used were sufficient to help the models recognize the main differences between the two general lithofacies (Fig. 15) recognized by previous works. Interestingly, for the third set processed, both algorithms confused the petrofacies YLLCON (concretions) and YLLCV (convoluted laminations) from the G1 group with the GLL petrofacies of the G2 group (Fig. 15). These petrofacies are linked to the increase in the mechanical strength of the strata due to early silica cementation (Figs. 2 and 3). The basal section of PW3 presented higher values of UCS, which is linked to G2 petrofacies in PW1 and PW2. As the SVM used the USC feature — unlike GNB — the SVM possibly tended to choose the "generic" GLL petrofacies because it had a higher frequency than the G2 group in general. This correlation between the mineralogical composition and the variation in the mechanical strength hindered the classification for part of the PW3 succession, which explains why this model performed worst on the third classification set.

An analysis of the efficiency of the models in the classification of the two groups of petrofacies was also considered because, in terms of reservoir modeling, a coarse scale of parameterization can

also be considered, as proposed in previous literature about these deposits. The models showed similar performances in interpreting the original proportions of groups G1 and G2 (Table V) for the first set in terms of percentage changes (12% for the G1 group and 9% for G2), with the best GNB model tending to overestimate the G1 group and underestimate G2, and the SVM presenting opposite results. Through the second set used, the best GNB model underestimated G1 by 12% and overestimated G2 by 38%, while the SVM reached a similar pattern with 18% and 58%, respectively. For the third set analyzed, the proportions of each group were extremely unbalanced, with 96% for G1 and 4% for G2. In this case, the best GNB model found a proportion of 61% and 39% for G1 and G2, and the SVM model found 52% and 48%, respectively. Consequently, the SVM model presented the lower performance.

The cross-validation and classification sets produced generally similar results, and also better achievements in terms of scoring metrics (especially in the case of the SVM model), when compared with other studies. Dunham et al. (2020) conducted an experiment on classifying 9 lithofacies in a scarce-data scenario (439 points with a sampling interval of 150cm), comparing GNB and SVM trained with semi-supervised approaches. The study demonstrated that the best accuracy results in cross-validation (5-fold) were 49.21% (GNB) and 50.41% (SVM), which are below the results presented in the present work. The results obtained by Silva et al. (2020) show that GNB had an accuracy slightly higher than 80%. However, the amount of available data (1,477 samples), the number of facies to be predicted (3), the application of previous preprocessing (SMOTE technique), and the method (using only cross-validation, without employing a blind test in a real well) used can affect a more accurate and unbiased diagnosis of model performance and make this experiment much less challenging than the present work. The same arguments can be extended to the work conducted by López & Thomas (2009).

CONCLUSIONS

We used ML models (GNB and SVM) to automatically classify petrofacies in three pseudo-wells in a thin laminated limestone unit (~20m thickness) with kilometers of lateral continuity, considering a reservoir cell scale (1 km between wells) and a limited number of logs (radiometry and mechanical strength). Both models successfully classified eight petrofacies for the wells PW1 and PW2 separated by 9.95 m. But, the classification for the third pseudo-well PW3, located 1 km away, was less effective. However, despite the challenges faced, the models presented relative success in defining the proportions of the two general lithological groups.

The results show that the selection of features used to build the GNB model was fundamental for maximizing its performance, whereas the SVM model performed best when using all available features. Thus, for the individual classification of petrofacies, the SVM model showed the best overall performance in the three configuration sets proposed. However, the GNB model performed better when analyzing the classification capacity for the two general groups of petrofacies (G1 and G2).

The study showed the main problems treating the classification of fine vertical variations in thin vertical intervals using the chosen ML models. It can help future understanding of how the increasing complexity of facies description can affect classification using ML in extreme scenarios such as the one treated here.

Acknowledgments

REFERENCES

ADAMS E, GRÉLAUD C, PAL M, CSOMA A, AL JA'AIDI O & HINAI R. 2011. Improving reservoir models of Cretaceous carbonates with digital outcrop modelling (Jabal Madmar, Oman): Static modelling and simulating clinoforms. Pet Geosci 17: 309-332. https://doi.org/10.1144/1354-079310-031.

AL-ANAZI A & GATES ID. 2010. A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. Eng Geol 114(3-4): 267-277. ISSN 0013-7952. https://doi.org/10.1016/j.enggeo.2010.05.005.

ALEXSANDRO GC, CARLOS ACP & GERALDO GN. 2017. Facies classification in well logs of the Namorado oilfield using Support Vector Machine algorithm. SEG Global Meeting Abstracts, 1853-1858. https://doi.org/10.1190/sbgf2017-365.

ANNA X, YUEXIA D, DAKUANG H, DAKUANG H, FENGCANG Y, CHUNLIANG H, ZECHENG W & XIAOFENG D. 2009. Integrated description and evaluation of reservoirs based on seismic, logging, and geological data: Taking Dongying Formation Member 1 oil reservoir of No. 1 structure, Nanpu Sag as an example. Pet Explor Dev 36(10): 541-551. https://10.1016/S1876-3804(09)60146-4.

ARAUJO A, RAMOS G, BARBOSA J, CORREIA FILHO O, ALENCAR M, MIRANDA T, NEUMANN V, CELESTINO AM & TOPAN JG. 2020. Depositional and diagenetic controls on the mechanical stratigraphy of Aptian laminated limestones of Crato Formation. Rio Oil Gas Expo Conf 20(12). https://doi.org/10.48072/2525-7579.rog.2020.035.

ASSINE M, PERINOTTO A, CUSTÓDIO MA, NEUMANN V, VAREJÃO FG & MESCOLOTTI PC. 2014. Sequências Deposicionais do Andar Alagoas (Aptiano superior) da Bacia do Araripe, Nordeste do Brasil. Bol Geocienc - Petrobras 22(06): 3-28.

BABASAFARI AA, VIDAL AC, CHINELATTO GF, GAVIDIA JCR & MATEUS B. 2022. Ensemble-based machine learning application for lithofacies classification in a pre-salt carbonate reservoir, Santos Basin, Brazil. Pet Sci Technol 1-17. doi: https://10.1080/10916466.2022.2143813.

BAYER P, COMUNIAN A, HÖYNG D & MARIETHOZ G. 2015. High resolution multi-facies realizations of sedimentary reservoir and aquifer analogs. Sci Data 2: 150033. https://doi.org/10.1038/sdata.2015.33.

BELAYNEH M, GEIGER S & MATTHAI S. 2006. Numerical simulation of water injection into layered fractured carbonate reservoir analogs. AAPG Bull 90(10): 1473-1493. https://doi.org/10.1306/05090605153.

BHATTACHARYA S, DOVETON JH, CARR T, GUY WR & GERLACH PM. 2005. Integrated core-log petrofacies analysis in the construction of a reservoir geomodel: A case study of a mature Mississippian carbonate reservoir using limited data. AAPG Bull 89(10): 1257-1274. https://10.1306/06030504144.

BISHOP CM. 2006. Pattern Recognition and Machine Learning. Information Science and Statistics. Springer, New York, NY, ISBN 978-0-387-31073-2, XX, 778 p.

BONNELL B & HURICH C. 2008. Characterization of Reservoir Heterogeneity: An Investigation of the Role of Cross-Well Reflection Data. Refl Dat: CSEG Rec 33(02).

BURCHETTE T. 2012. Carbonate rocks and petroleum reservoirs: A geological perspective from the industry. Geol Soc Lond Spec Publ 370(11): 17-37. https://doi.org/10.1144/SP370.14.

CAO B, LUO X, ZHANG L, LEI Y & ZHOU J. 2020. Petrofacies prediction and 3-D geological model in tight gas sandstone reservoirs by integration of well logs and geostatistical modeling. Mar Pet Geol 114(12): 104202. https://doi.org/10.1016/j.marpetgeo.2019.104202.

CAO B, SUN W & LI J. 2021. Reservoir petrofacies — A tool for characterization of reservoir quality and pore structures in a tight sandstone reservoir: A study from the sixth member of Upper Triassic Yanchang Formation, Ordos Basin, China. J Pet Sci Eng 199(12): 108294. https://doi.org/10.1016/j.petrol.2020.108294.

CATTO B, JAHNERT RJ, WARREN LV, VAREJAO FG & ASSINE ML. 2016. The microbial nature of laminated limestones: Lessons from the Upper Aptian, Araripe Basin, Brazil. Sediment Geol 341(05): 304-315. https://doi.org/10.1016/j.sedgeo.2016.05.007.

CELESTINO M, MIRANDA TS, MARIANO G, ALENCAR ML, CARVALHO BRBM, FALCÃO TC, TOPAN JG, BARBOSA JS & GOMES IG. 2020. Fault damage zones width: Implications for the tectonic evolution of the northern border of the Araripe Basin, Brazil, NE Brazil. J Struct Geol 138(06): 104116. https://doi.org/10.1016/j.jsg.2020.104116.

CHAWSHIN K, BERG CF, VARAGNOLO D & LOPEZ O. 2021. Lithology classification of whole core CT scans using

convolutional neural networks. SN Appl Sci 3(06). https://doi.org/10.1007/s42452-021-04656-8.

CORREIA U, BATEZELLI A & LEITE E. 2016. 3-D Geological modelling: a siliciclastic reservoir case study from Campos Basin, Brazil. Rem Rev Esc Minas 69(10): 409-416. https://doi.org/10.1590/0370-44672015690063.

CORTES C & VAPNIK V. 1995. Support Vector Network. Machine Learn 20(09): 273-297. https://doi.org/10.1007/BF00994018.

CREUSEN A, MAARI KK, TULL S, VAHRENKAMP VC, MOOKERJEE A & VAN RIJEN M. 2007. Property Modelling Small Scale Heterogeneity of Carbonate Facies. Paper presented at the SPE/EAGE Reservoir Characterization and Simulation Conference, Abu Dhabi, UAE. https://doi.org/10.2118/111451-MS.

DENG T, XU C, JOBE D & XU R. 2019. A Comparative Study of Three Supervised Machine-Learning Algorithms for Classifying Carbonate Vuggy Facies in the Kansas Arbuckle Formation. Petrophysics 60: 838-853. doi: https://doi.org/10.30632/PJV60N6-2019a8.

DUARTE D, LIMA RP, TELLEZ & PRANTER MJ. 2023. Spatial variability of petrofacies using supervised machine learning and geostatistical modeling: Sycamore Formation, Sho-Vel-Tum Field, Oklahoma, USA. Interpretation 11: 1-44. https://10.1190/int-2022-0064.1.

DUNHAM M, MALCOLM A & WELFORD K. 2020. Improved Well-log classification using semisupervised label propagation and self-training, with comparisons to popular supervised algorithms. Geophysics 85(01): 1-15. https://doi.org/10.1190/geo2019-0238.1.

EDWARDS J, LALLIER F, CAUMON G & CARPENTIER C. 2017. Uncertainty management in stratigraphic well correlation and stratigraphic architectures: A training-based method. Comput Geosci 111(10). https://10.1016/j.cageo.2017.10.008.

ENGE H, BUCKLEY SJ, ROTEVATN & HOWELL J. 2007. From Outcrop to reservoir simulation model: Workflow and procedures. Geosphere 3(6): 469-490. https://doi.org/10.1130/GES00099.1.

ERTEKIN T & SUN Q. 2019. Artificial Intelligence Applications in Reservoir Engineering: A Status Check. Energies 12(07): 2897. https://doi.org/10.3390/en12152897.

FADOKUN DO, OSHILIKE IB & MIKE OO. 2020. Supervised and Unsupervised Machine Learning Approach in Facies Prediction. Paper presented at the SPE Nigeria Annual International Conference and Exhibition, Virtual. doi: https://doi.org/10.2118/203726-MS.

FALIVENE O, ARBUÉSP, HOWELL J, MUÑOZ JA, FERNÁNDEZ O & MARZO M. 2006. Hierarchical geocellular facies modelling of a turbidite reservoir analogue from the Eocene of the Ainsa basin, NE Spain. Mar Pet Geol 23(06): 679-701. https://doi.org/10.1016/j.marpetgeo.2006.05.004.

GÓMEZ JJ, CARREÑO AM, PÉREZ MR, PEREZ ER, PEÑA LF, RUEDA LH, RÍOS CA & HENAO JA. 2020. Petrofacies and diagenetic processes of la Victoria Formation (Early Miocene), Dina oil field, upper Magdalena valley basin, Colombia. CT y F - Cienc Tecnol Futuro 10(1): 33-44. https://doi.org/10.29047/01225383.120.

GONZALEZ A, HEIDARI Z & OLIVIER L. 2023. Data-Driven Algorithms for Image-Based Rock Classification and Formation Evaluation in Formations With Rapid Spatial Variation in Rock Fabric. Petrophysics 64: 154-173. doi: https://doi.org/10.30632/PJV64N2-2023a2.

GRANA D, LANG X & WU W. 2016. Statistical facies classification from multiple seismic attributes: comparison between Bayesian classification and expectation-maximization method and application in petrophysical inversion: Statistical facies classification. Geophys Prospect 65(08). https://doi.org/10.1111/1365-2478.12428.

HAESE R, JIANG G, OOI S & BLACK JR. 2019. Enhanced Reservoir Characterization Using Hyperspectral Core Logging. J Sediment Res 89(03): 187-198. https://doi.org/10.2110/jsr.2019.10.

HALL B. 2016. Facies classification using machine learning. Lead Edge 35(10): 906-909. https://doi.org/10.1190/tle35100906.1.

HALOTEL J, DEMYANOV V & GARDINER A. 2019. Value of Geologically Derived Features in Machine Learning Facies Classification. Math Geosci 52: 5-29. https://doi.org/10.1007/s11004-019-09838-0.

HEIDSIEK M, BUTSCHER C, BLUM P & FISCHER C. 2020. Small-scale diagenetic facies heterogeneity controls porosity and permeability pat-tern in reservoir sandstones. Environ Earth Sci 79(09): 25. https://doi.org/10.1007/s12665-020-09168-z.

HEIMHOFER U, ARIZTEGUI D, LENNIGER M, HESSELBO S, MARTILL D & RIOS-NETTO A. 2010. Deciphering the depositional environment of the laminated Crato fossil beds (Early Cretaceous, Araripe Basin, North eastern Brazil). Sedimentology 57(11): 677-694. https://doi.org/10.1111/j.1365-3091.2009.01114.x.

HORROCKS T, HOLDEN E & WEDGE D. 2015. Evaluation of automated lithology classification architectures using highly-sampled wireline logs for coal exploration. Comput Geosci 83: 209-218. ISSN 0098-3004. https://doi.org/10.1016/j.cageo.2015.07.013.

HOWELL J, MARTINIUS A & GOOD T. 2014. The application of outcrop analogues in geological modelling: a review, present status and future outlook. Geol Soc Lond Spec Publ 387(06): 1-25. https://doi.org/10.1144/SP387.12.

JARDIM C, DE ROS L & KETZER M. 2011. Reservoir quality assessment and petrofacies of the Lower Cretaceous siliciclastic, carbonate and hybrid arenites from the Jequitinhonha Basin, Eastern Brazil. J Pet Geol 34(07). https://doi.org/10.1111/j.1747-5457.2011.00507.x.

JONES RR, MCCAFFREY KJW, IMBER J & WIGHTMAN R. 2008. Calibration and validation of reservoir models: the importance of high resolution, quantitative outcrop analogues. Geol Soc Lond Spec Publ 309: 87-98. https://doi.org/10.1144/SP309.7.

KADKHODAIE A & KADKHODAIE-ILKHCHI R. 2018. A Review of Reservoir Rock Typing Methods in Carbonate Reservoirs: Relation between Geological, Seismic, and Reservoir Rock Types. Iran J Oil Gas Sci Technol 7(12): 13-35. https://doi.org/10.22050/IJOGST.2019.136243.1461.

KUMAR T, SEELAM NK & RAO GS. 2022. Lithology prediction from well log data using machine learning techniques: A case study from Talcher coalfield, Eastern India. J Appl Geophys 199: 104605. ISSN 0926-9851. https://doi.org/10.1016/j.jappgeo.2022.104605.

LAUBACH SE ET AL. 2019. The Role of Chemistry in Fracture Pattern Development and Opportunities to Advance Interpretations of Geological Materials. Rev Geophys 57(08). https://doi.org/10.1029/2019RG000671.

LIKUAN Z ET AL. 2021. Small-Scale Diagenetic Heterogeneity Effects on Reservoir Quality of Deep Sandstones: A Case Study from the Lower Jurassic Ahe Formation, Eastern Kuqa Depression. Geofluids 2021(03): 1-25. https://10.1155/2021/6626652.

LIMA R, SURIANAM F, MARFURT K & PRANTER MJ. 2019. Convolutional neural networks as aid in core lithofacies classification. Interpretation 7(03): 1-50. https://doi.org/10.1190/INT-2018-0245.1.

LINEMAN DJ, MENDELSON JD & TOKSOZ MN. 1987. Well-to-Well Log Correlation Using Knowledge-Based Systems and Dynamic Depth Warping. Massachusetts Institute of Technology, Earth Resources Laboratory, vol. 86.

LI Y & ANDERSON-SPRECHER R. 2006. Facies identification from well logs: A comparison of discriminant analysis and Naïve Bayes classifier. J Pet Sci Eng 53(3-4): 149-157. DOI: https://10.1016/j.petrol.2006.06.001.

LIU X Y, ZHOU L & CHEN X H. 2020. Lithofacies identification using support vector machine based on local deep multi-kernel learning. Pet Sci 17: 954-966. https://doi.org/10.1007/s12182-020-00474-6.

LÓPEZ CC & THOMAS D. 2009. Sandstone Petrofacies Prediction to Characterize Permeability For Postle Field, Oklahoma. Paper presented at the 2009 SEG Annual Meeting, Houston, Texas.

MEREMBAYEV T, KURMANGALIYEV D, BEKBAU B & AMANBEK Y. 2021. A Comparison of Machine Learning Algorithms in Predicting Lithofacies: Case Studies from Norway and Kazakhstan. Energies 14(03): 1896. https://doi.org/10.3390/en14071896.

MIKES D, BARZANDJI OHM, BRUINING J & GEEL CR. 2006. Upscaling of small-scale heterogeneities to flow units for reservoir modelling. Mar Pet Geol 23(9-10): 931-942. https://doi.org/10.1016/j.marpetgeo.2005.06.005.

MILAD B & SLATT R. 2019. Outcrop to Subsurface Reservoir Characterization of the Mississippian Sycamore/Meramec Play in the SCOOP Area, Arbuckle Mountains, Oklahoma, USA. Paper presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference, Denver, Colorado, USA. https://doi.org/10.15530/urtec-2019-991.

MIRANDA TS, SANTOS RF, BARBISA JA, GOMES IF, ALENCAR ML, CORREIA OJ, FALCÃO TC, GALE JFW & NEUMANN VH. 2018. Quantifying aperture, spacing and fracture intensity in a carbonate reservoir analogue: Crato Formation, NE Brazil. Mar Pet Geol 97(08): 556-567. https://doi.org/10.1016/j.marpetgeo.2018.07.019.

MURPHY KP. 2012. Machine Learning: A Probabilistic Perspective. Illustrated edition. The MIT Press. ISBN 978-0262018029, 1104 p.

NEUMANN V, BORREGO AG, CABRERA L & DINO R. 2003. Organic matter composition and distribution through the Aptian–Albian lacustrine sequences of the Araripe Basin, northeastern Brazil. Int J Coal Geol 54(04): 21-40. https://doi.org/10.1016/S0166-5162(03)00018-1.

NEUMANN V, CABRERA L & CALVET F. 1999. Estratigrafia, sedimentologia, geoquimica y diagenesis de los sistemas lacustres aptiense-albienses de la cuenca de araripe(noreste de brasil). Ph.D. thesis, Universitat de Barcelona.

NGUYEN NT, HUNG T, HOANG S & TUNG T. 2022. Supervised machine learning application of lithofacies classification for a hydrodynamically complex gas - condensate reservoir in Nam Con Son basin. Petrovietnam J 6: 27-35. https://doi.org/10.47800/PVJ.2022.06-03.

NWAILA GT, ZHANG SE & BOURDEAU JE. 2022. Data-Driven Predictive Modeling of Lithofacies and Fe In-Situ Grade

in the Assen Fe Ore Deposit of the Transvaal Supergroup (South Africa) and Implications on the Genesis of Banded Iron Formations. Nat Resour Res 31: 2369-2395. https://doi.org/10.1007/s11053-022-10105-z.

OSÉS G ET AL. 2017. Deciphering pyritization-kerogenization gradient for fish soft-tissue preservation. Sci Rep: 7(12). https://doi.org/10.1038/s41598-017-01563-0.

OTHMAN A, MUHAMMAD M & MOHAMED I. 2021. Application of Artificial Neural Network in seismic reservoir characterization: a case study from Offshore Nile Delta. Earth Sci Inform 14(06). https://10.1007/s12145-021-00573-x.

OZKAN A, CUMELLA S, MILLIKEN K & LAUBACH S. 2011. Prediction of lithofacies and reservoir quality using well logs, Late Cretaceous Williams Fork Formation, Mamm Creek field, Piceance Basin, Colorado. AAPG Bull 95(10): 1699-1723. https://doi.org/10.1306/01191109143.

QUESTIAUX JM, COUPLES G & RUBY N. 2009. Fractured Reservoirs with fracture corridors. Geophys Prospect 58(05): 279-295. https://doi.org/10.1111/j.1365-2478.2009.00810.x.

RAEESI M, MORADZADEH A, ARDEJANI F & RAHIMI M. 2012. Classification and identification of hydrocarbon reservoir lithofacies and their heterogeneity using seismic attributes, logs data and artificial neural networks. J Pet Sci Eng 82-83(02): 151-165. https://doi.org/10.1016/j.petrol.2012.01.012.

RAMOS G, ARAUJO A, BARBOSA J, CELESTINO MAL, CORREIA O, ALENCAR ML, MIRANDA TS, MIRANDA VH & GABRIEL J. 2020. Estimation of compaction rates for a tight carbonate reservoir based in analyses of diagenetic concretions. Rio Oil Gas Expo Conf 20(12). https://doi.org/10.48072/2525-7579.rog.2020.036.

RINGROSE P, MARTINIUS A & ALVESTAD J. 2008. Multiscale geological reservoir modelling in practice. Geol Soc Lond Spec Publ 309(01): 123-134. https://doi.org/10.1144/SP309.9.

SANTOS R, TIAGO M, BARBOSA A, GOMES I, MATOS G, GALE JFW, NEUMANN V & GUIMARÃES LJN. 2015. Characterization of natural fracture systems: Analysis of uncertainty effects in linear scanline results. AAPG Bull 99(12): 2203-2219. https://doi.org/10.1306/05211514104.

SARKAR S & MAJUMDAR C. 2020. A Comparative Analysis of Supervised Classification Algorithms for Lithofacies Characterization. Eur Assoc Geosci Eng, First EAGE Digitalization Conf Exhib, 2020, p. 1-5. doi: https://doi.org/10.3997/2214-4609.202032090.

SILVA A ET AL. 2020. Petrofacies classification using machine learning algorithms. Geophysics 85(03) 1-71. https://doi.org/10.1190/geo2019-0439.1.

THOMAS A, RIDER M, CURTIS A & MACARTHUR A. 2011. Automated lithology extraction from core photographs. First Break 29(06). https://doi.org/10.3997/1365-2397.29.6.51281.

VAPNIK V. 1995. The Nature Of Statistical Learning Theory. Springer, vol. 6.

VAPNIK V & CHERVONENKIS A. 1971. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. Theory Probab Its Appl 17(01): 264-280. https://doi.org/10.1137/1116025.

VERMA S, BHATTACHARYA S, CHOWDHURY NUMD & TIAN M. 2021. A new workflow for multi-well lithofacies interpretation integrating joint petrophysical inversion, unsupervised, and supervised machine learning. Paper presented at the SEG/AAPG/SEPM First International Meeting for Applied Geoscience & Energy, Denver, Colorado, USA and online. doi: https://doi.org/10.1190/segam2021-3584118.1.

WATNEY W, BHATTACHARYA S, CHOWDHURY NUMK & TIAN M. 1998. Petrofacies Analysis - A Petrophysical Tool for Geologic/Engineering Reservoir Characterization. USGS Publications Warehouse 71: 73-90. http://pubs.er.usgs.gov/publication/70019834.

WITTEN I & FRANK E. 2002. Data Mining: Practical MachineLearning Tools and Techniques with Java Implementations. SIGMOD Rec 31(1): 76-77. https://doi.org/10.1145/507338.507355.

WORDEN R, OMMA J, HOLLIS C, LANDER RH, CSOMA AÉ, BUTCHER AR & ARMITAGE PJ. 2018. Petroleum reservoir quality prediction: Overview and contrasting approaches from sandstone and carbonate communities. Geol Soc Lond Spec Publ 435(05): SP435.21. https://doi.org/10.1144/SP435.21.

WU X, SHI Y, FOMEL S & LI F. 2018. Incremental correlation of multiple well logs following geologically optimal neighbors. Interpretation 6(06): T713-T722. https://10.1190/int-2018-0020.1.

XIE Y, ZHU C, ZHOU W, LI Z, LIU X & TU M. 2018. Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. J Pet Sci Eng 160: 182-193. https://doi.org/10.1016/j.petrol.2017.10.028.

YAN Y, LIQIANG Z & LUO X. 2020. Modeling Three-Dimensional Anisotropic Structures of Reservoir Lithofacies Using Two-Dimensional Digital Outcrops. Energies 13(16): 4082. https://doi.org/10.3390/en13164082.

YU X, LI S, ZHAO S, CHEN J & HOU G. 2008. Constraining Method of Stochastic Modeling for Fluvial Petroleum Reservoir Controlled by Depositional Facies Using Wells and Seismic Data. Earth Sci Front 15(07): 33-41. https://10.1016/S1872-5791(08)60037-8.

ZHANG T, BOMBERDE S, STREBELLE S & OATNEY E. 2006. 3D Porosity Modeling of a Carbonate Reservoir Using Continuous Multiple Point Statistics Simulation. SPE J 11(09): 375-379. https://doi.org/10.2118/96308-PA.

ZHAO L, GENG J, CHENG J, HAN D-H & TONGLOU G. 2014. Probabilistic lithofacies prediction from prestack seismic data in a heterogeneous carbonate reservoir. Geophysics 79(08): M25-M34. https://doi.org/10.1190/geo2013-0406.1.

How to cite

GALLILEU GENESIS[1]
https://orcid.org/0000-0003-2452-2076

IGOR F. GOMES[1]
https://orcid.org/0000-0003-2474-383X

JOSÉ ANTONIO BARBOSA[2]
https://orcid.org/0000-0001-8754-6310

ARALY FABIANA L. DE ARAÚJO[2]
https://orcid.org/0000-0002-8078-0670

GERMANO MÁRIO S. RAMOS[2]
https://orcid.org/0000-0003-0991-8929

[1]Universidade Federal de Pernambuco, Departamento de Engenharia Civil, Laboratório de Métodos Computacionais em Geomecânica, Av. da Arquitetura, s/n, Cidade Universitária, 50740-540 Recife, PE, Brazil

[2]Universidade Federal de Pernambuco, Departamento de Geologia, GEOQUANTT, Av. da Arquitetura, Cidade Universitária, 57040-530 Recife, PE, Brazil

Correspondence to: **Gallileu Genesis**
E-mail: galileugenesis@gmail.com

Author contributions

G.G, I.F.G. and J.A.B. conceived the idea and the contribution potential. A.F.L.A. and G.M.R.S. processed the field data and performed the analysis focusing on constructing the pseudo-wells and the petrofacies characterization. G.G. performed the computational analysis and the results. G.G., J.A.B., IFG, A.F.L.A. and G.M.S.R. prepared the manuscript and figures. G.G, I.F.G. and J.A.B. discussed the results and commented on the final version of the manuscript.