ISSN 1993-9973, e-ISSN 2415-332X

Vol. 26, No. 1
2026

PROSPECTING AND DEVELOPMENT OF OIL AND GAS FIELDS

https://pdogf.com.ua/en

Received: 27.10.2025. Revised: 20.04.2026. Accepted: 29.05.2026. Published: 29.06.2026.

UDC 550.834:004.85

DOI: 10.63341/pdogf/1.2026.16

Machine learning-enhanced seismic attribute analysis for reservoir quality prediction in clastic oil fields

Fahri Maho*

PhD in Water Engineering, Lecturer
Polytechnic University of Tirana
1001, 28 Teodor Keko Str., Tirana, Albania
https://orcid.org/0009-0002-0673-4577

Enkelejda Gjinali

PhD in Water Environmental Engineering, Associate Professor
Polytechnic University of Tirana
1001, 28 Teodor Keko Str., Tirana, Albania
https://orcid.org/0009-0005-4942-3893

Abstract. The aim of the study was to assess machine learning-enhanced seismic attribute analysis for reservoir quality prediction in clastic and Balkan-type oil fields. The validation study was performed on the open-access Project Penobscot seismic and reservoir data for a range of attributes within the predictor space, as well as for the four reservoir quality variables. XGBoost showed the best regression results for all continuous parameters, with the highest accuracy for porosity ($R^2 = 0.88$, root mean square error = 0.024, mean absolute error = 0.018), followed by the sandiness index ($R^2 = 0.85$, root mean square error = 0.054, mean absolute error = 0.039), whereas permeability remained the most difficult parameter to predict ($R^2 = 0.76$, root mean square error = 16.9, mean absolute error = 12.4). Facies classification also demonstrated stable performance, with accuracy of 0.84, precision of 0.82, recall of 0.79, and F1-score of 0.8. The predicted parameter distributions revealed a heterogeneous reservoir system, with porosity as the most stable parameter, sandiness index showing intermediate variability, and permeability displaying the highest contrast. Integrated interpretation of the predicted indicators differentiated the reservoir into high-, moderate-, and low-quality zones with indices of 86, 64, and 38, respectively. In conclusion, reservoir quality is primarily controlled by the factors related to storage and flow potential, sand bodies, and facies, yet machine learning-enhanced methods can effectively assist reservoir engineers in making accurate reservoir predictions in these heterogeneous systems. The practical value of this study is in the potential use of this established workflow to effectively screen for and prioritise oil and gas production within the reservoirs with the help of machine learning, especially those geologically structurally complex clastic systems

Keywords: porosity; square error; permeability; flow potential; sandiness index; Balkan-type

Introduction

Due to the growing complexity of the reservoir prediction in clastic oil fields, there is an increasing interest in utilising data-driven approaches for the interpretation of seismic information. In many instances, particularly in those reservoirs with high degrees of structural complexity, the seismic data can be utilised to make predictions regarding the properties of the reservoir that is being evaluated. Moreover, many of these reservoirs are among those that are considered to be mature or have developed structural complexities that make it particularly difficult to directly observe the reservoirs. Thus, any method that can help to extract information regarding those reservoirs is of interest to those within the oil industry. One such approach that has developed in response to these needs is the application of machine learning approaches to the analysis of seismic data to help extract information regarding the reservoirs.

Suggested Citation: Maho, F., & Gjinali, E. (2026). Machine learning-enhanced seismic attribute analysis for reservoir quality prediction in clastic oil fields. *Prospecting and Development of Oil and Gas Fields*, 26(1), 16-29. doi: 10.63341/pdogf/1.2026.16.

*Corresponding author (*mahofahri2@gmail.com*)

For the Balkan region, specifically for Albania, this problem has certain geological and applied significance. The segment of the Ionian Zone that lies within Albania is tectonically complex, has varying strata within its geologic outcrops, and has a general heterogeneity of its sedimentary features, all of which make it difficult for those attempting to use seismic surveys to model the region to predict the reservoir quality within the area. Petrographic and mineralogical research of the Oligocene flysch of the Ionian Zone of Albania, performed by A. Fociro *et al*. (2023), has determined the compositional heterogeneity of the clastic strata of the region, indicating the importance of its sedimentary origin for explaining the geological heter ogeneity of the area. In the work of J.I. Soto *et al*. (2024) showed that the Ionian Zone of Greece and Albania underwent contrasting types of salt-tectonic processes that significantly affected the structural configuration, trap architecture, and geometry of subsurface bodies. Such features are of fundamental importance for seismic interpretation, since deformation, changes in thickness of strata, and structural blockiness directly change the wave pattern and spatial continuity of reservoirs. An additional regional context was provided by the study of I. Alexandridis *et al*. (2022), in which a new interval of parent rocks was discovered in the adjacent segment of the Ionian zone in the west of Greece, which once again emphasises the oil and gas potential of this tectonic-sedimentary complex for the entire Balkan region. Collectively, these works showed that the Balkan oil and gas context needs such interpretive approaches that are able to combine lithological variability, structural deformation and indirect prediction of petrophysical properties in a single analytical framework. At the same time, modern international studies have proven that machine learning significantly expands the analytical possibilities of interpreting seismic attributes in tasks of reservoir characterisation (Oitseva, 2024). In the work of D.J. Ferreira *et al*. (2021) the classification of seismic patterns was integrated with porosity and permeability estimation, which showed the possibility of direct connection of multidimensional seismic analysis with the prediction of reservoir properties, and not only with its descriptive interpretation. M.H. Rahma Putra *et al*. (2024) confirmed that Gaussian process machine learning can be effectively used for porosity estimation and anomaly detection based on seismic attributes, increasing prediction reliability with limited direct control. In the work of Y. Chen *et al*. (2021) applied machine learning with physical constraints and multiple seismic attributes for deep reservoir characterisation, which is methodologically important because it demonstrates the need to combine predictive flexibility of algorithms with geological plausibility of results. This aspect is especially important in the case of clastic reservoirs, as the purely algorithmic recognition of patterns within seismic data without geological controls may lead to unstable reservoir predictions.

Additional evidence of the effectiveness of machine learning approaches to reservoir description is provided in the article by S. Narayan *et al*. (2024). In this article, the authors discuss the integration of supervised machine learning techniques with seismic impedance inversion methods. The importance of this method is that it allows for the inverted seismic data and attributes to be considered as potential predictors of reservoir characteristics, in addition to the traditional methods of reservoir modelling. Additionally, S.A. El-Dabaa *et al*. (2024) applied unsupervised machine learning techniques to reservoir data with the goal of determining the importance of the different dimensions of seismic data in the prediction of reservoir quality. Author M. Ali *et al*. (2024) described the use of synthetic data and machine learning to model reservoir characteristics from seismic data alone, indicating that machine learning methods have good generalisation abilities. Furthermore, Z.A. Riyadi *et al*. (2024) utilised machine learning to model the prediction of reservoir permeability from elastic properties, which are factors in the determination of the reservoir quality. From the point of view of Water Engineering, reservoir quality is a factor in the determination of the reservoir's ability to accumulate and move fluids.

These methods can be particularly valuable in areas like Albania and the Balkan region in particular, where structural geology makes it difficult to determine reservoir quality. In these cases, machine learning methods can help to improve the robustness of reservoir forecasting techniques by considering various seismic data descriptors of the reservoir. Furthermore, the use of such methods should be based upon the geological context of the reservoir, which can include factors like sedimentation conditions, the blockiness of the structure, or the way in which the predicted properties may impact the movement of fluids within the reservoir media. Therefore, the research is aimed at forming a unified analytical approach to the analysis of seismic attributes, enhanced by machine learning, for predicting reservoir quality in clastic oil deposits, taking into account the specifics of the Balkan region and Albania. The work combines seismic characterisation of the reservoir, oil and gas geology and interpretation of the filtration properties of the porous medium.

The aim of this study was to determine the potential of the use of machine learning enhanced seismic attribute analysis for the prediction of reservoir quality in clastic oil fields, specifically within the Balkan region and the country of Albania. The specific objectives of this study included the survey of the geological and geodynamic factors related to the prediction of reservoir quality in the Balkans region; the evaluation of the various machine learning methods that can be used to predict reservoir quality parameters; and the development and testing of a methodology that employs seismic attributes and machine learning methods to predict reservoir quality and fluid-flow potential in clastic reservoirs.

Materials and Methods

The study was conducted using the Project Penobscot dataset to predict reservoir quality in clastic oil fields based on seismic attributes and machine learning algorithms, with interpretive relevance for Albanian and other Balkan-type reservoir systems. The empirical basis of the study was the open-access Project Penobscot dataset, which was selected as a publicly available geophysical test site suitable for validating an integrated workflow for seismic attribute-based reservoir quality prediction. Open-access regional subsurface datasets directly comparable to Project Penobscot are currently limited for Albania and the wider Balkan petroleum domain; therefore, the Penobscot dataset was used as an empirical benchmark for workflow validation, while the geological applicability of the approach was interpreted with reference to clastic reservoir systems relevant to the Balkan region and Albania.

The dataset used for this study was obtained from the publicly available Project Penobscot project offshore Nova Scotia, Canada. The material base primarily included a 3D post-stack seismic volume $S(x, y, t)$, interpreted stratigraphic horizons H_i, stratigraphic markers, and well-log data $L(z)$, which formed the core basis for seismic attribute extraction and target calibration. Additional dataset components, including 2D seismic profiles and prestack information, were considered only as auxiliary support during interval familiarisation and verification and were not used as an independent basis for the reported predictive results. The 3D seismic cube $S(x, y, t)$ served as the main source for seismic attribute extraction, whereas the well-log data $L(z)$ served as the reference source for reservoir-related target parameters. The interpreted horizons H_i and stratigraphic markers were used to constrain the analysis to geologically consistent intervals and to maintain a stable relationship between seismic reflections and reservoir units.

A working multi-attribute set was constructed from the seismic volume to represent seismic descriptors specifically relevant to the geological setting of Albanian and other Balkan-type clastic reservoirs, where tectonic deformation, stratigraphic variability, internal heterogeneity, and discontinuous sand-body architecture affect reflection continuity, amplitude response, frequency behaviour, and impedance contrasts that are indirectly related to porosity, permeability, and facies organisation. This set included amplitude-related attributes A_{amp}, represented by instantaneous amplitude A_{inst} and RMS amplitude A_{RMS}; frequency-sensitive attributes A_{freq}, represented by instantaneous frequency f_{inst} and sweetness SW; geometric and continuity-related attributes A_{geom}, including coherence C and variance Var; and inversion-derived parameters A_{inv}, particularly acoustic impedance AI. The analytical dataset was represented by the predictor matrix (1):

$$X = [A_1, A_2, \ldots, A_n]. \tag{1}$$

The vector of continuous target variables (2):

$$Y_c = [\phi, k, SI]. \tag{2}$$

The categorical target variable used for facies classification (3):

$$Y_f = F, \tag{3}$$

where ϕ – porosity; k – permeability; SI – the sandiness index; F – the facies class. Porosity is an estimation based on well log data and is used as the main indicator of the reservoir's storage capacity. Permeability is obtained from the reference petrophysical dataset and represents the permeability of the selected well intervals. The sandiness index is calculated from the sandy nature of the intervals. The facies class is also determined from the well log data and divides the intervals into different facies according to their lithology. Thus, these parameters represent both the continuous regression targets (ϕ, k, SI) and the categorical classification target (F). These target variables were selected because, taken together, they provide an operational representation of reservoir quality in clastic systems and allow subsequent engineering interpretation in terms of storage capacity, filtration potential, internal heterogeneity, and probable fluid-flow behaviour.

The analytical sample was non-random and condition-based. It was formed only from intervals for which consistency between $S(x, y, t)$, H_i, and $L(z)$ could be ensured. The inclusion criteria required acceptable seismic signal quality, stable time-depth alignment, interpretable stratigraphic correlation, and sufficiently complete well-log coverage. Intervals affected by major noise contamination, incomplete logging curves, unstable stratigraphic attribution, or ambiguous seismic-to-well ties were excluded. Following data cleaning, interval alignment, and preparation of the predictor and target variables for inclusion into the analysis, the dataset comprised 324 interval-level observations from 2 wells (L-30 and B-41) from the Project Penobscot dataset. Each observation in the dataset represented one individual interval within the wells. This predictor-target matrix served as the basis for subsequent model training and validation.

Data pre-processing, seismic attribute extraction, model development, validation, and result analysis were carried out using OpendTect, Petrel, Python 3.11, OriginPro 2024, and Microsoft Excel 365. The methodological workflow consisted of five successive stages: data preparation, attribute processing, feature selection, model training and validation, and integrated interpretation of predicted reservoir-quality indicators. At the first stage, seismic and well data were checked for consistency within a common interpretation framework. Seismic reflections were tied to the available well-log intervals using interpreted horizons and stratigraphic markers in order to ensure that predictor variables and target parameters referred to the same geological intervals. Only those intervals for which the time-depth relationship was sufficiently stable were retained for analysis. At the second stage, all continuous input variables were normalised using z-score transformation (4):

$$Z_i = \frac{X_i - \mu}{\sigma}, \tag{4}$$

where X_i – original feature value; μ – the mean; σ – standard deviation. To reduce multicollinearity within the predictor space, pairwise Pearson correlation screening was applied; variables with $|r| > 0.85$ were considered redundant. Highly correlated predictors were removed, after which recursive feature elimination was applied to retain a compact subset of informative variables for model training. At the third stage, the dataset was divided into training and testing subsets with an 80:20 ratio. The split was performed randomly at the interval level for the regression tasks and with stratification for the facies classification task in order to preserve class proportions. For regression tasks, Random Forest, XGBoost, and Gaussian Process Regression were used to predict continuous reservoir parameters. For facies classification, Support Vector Machine was applied. Model robustness and generalisation capability were assessed using 5-fold cross-validation. At the fourth stage, the predictive performance of regression models was evaluated using the coefficient of determination (5), root mean square error (RMSE) (6) and mean absolute error (MAE) (7):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}; \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}; \tag{6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |, \tag{7}$$

where y_i – reference value; $\hat{y}_i$ – predicted value; $\bar{y}$ – mean of reference values; n – the number of observations. For classification tasks, model quality was assessed using accuracy (8), precision (9), recall (10), and F1-score (11):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}; \tag{8}$$

$$Precision = \frac{TP}{TP+FP}; \tag{9}$$

$$Recall = \frac{TP}{TP+FN}; \tag{10}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}, \tag{11}$$

where *TP*, *TN*, *FP*, and *FN* denote true positive, true negative, false positive, and false negative predictions, respectively. In addition to mean, standard deviation, minimum, and maximum values, the coefficient of variation was calculated to compare the relative variability of the predicted reservoir parameters. The relative discriminative role of the continuous reservoir-quality parameters was interpreted on the basis of their normalised spread and coefficient of variation, with higher values indicating a stronger contribution to the differentiation of reservoir-quality groups. At the final stage, the trained models were applied to the attribute space to generate continuous and categorical prediction outputs. For integrated reservoir-quality interpretation, the predicted indicators were first transformed to a common normalised scale from 0 to 100. For the continuous variables, min-max normalisation was applied as follows (12):

$$V_i^{norm} = 100 \cdot \frac{V_i - V_{min}}{V_{max} - V_{min}}, \tag{12}$$

where V_i – predicted value of the corresponding parameter; V_{min}, V_{max} – the minimum and maximum values of that parameter within the analysed dataset. Facies class was incorporated as an ordinal categorical component reflecting the relative geological favourability of facies types for reservoir development. The integrated reservoir-quality index was then calculated as (13):

$$RQI = w_\phi \, \phi^{norm} + w_k \, k^{norm} + w_{SI} \, SI^{norm} + w_F \, \mathrm{F}^{norm}, \tag{13}$$

where w_ϕ, w_k, w_{SI}, and w_F – weights assigned to the normalised components. In the absence of field-specific prior weighting constraints, equal weights were used ($w_\phi = w_k = w_{SI} = w_F = 0.25$). The resulting index was used to differentiate the analysed intervals into high-, moderate-, and low-quality groups. For integrated reservoir-quality interpretation, the predicted indicators were transformed to a common comparative scale and combined into a composite ranking scheme. Porosity, permeability, and sandiness index contributed to the integrated assessment as continuous variables, whereas facies class was incorporated as a categorical factor providing geological context for zone differentiation. On this basis, the analysed reservoir intervals were classified into high-, moderate-, and low-quality groups. The results presented in the paper were summarised through comparative model-performance metrics, descriptive statistics of the predicted parameters, and integrated reservoir-quality ranking. The interpretation focused on reservoir heterogeneity, storage capacity, permeability distribution, internal compartmentalisation, and probable fluid-flow behaviour within clastic reservoir systems.

Results

Seismic attribute space as a predictor basis for reservoir quality analysis

The resulting predictor space *X*, as defined in Equation (1), indicates that the reservoir quality prediction in the clastic oil fields can be made possible by a set of seismic attributes with different functions. The seismic volume that was processed along with the seismic data provided a set of attributes that collectively reflected the seismic responses associated with the lithology of the subsurface formation. This shows that the predictor space was appropriately defined to include the attributes that would best represent the main features of the reservoir in terms of its impact upon reservoir quality. Furthermore, the obtained attributes indicated that the individual groups of selected attributes provided slightly different information from one another. The amplitude attributes were used to recognise the contrasts within the subsurface from the strength of the seismic reflections. The frequency attributes were used to recognise the contrasts within the subsurface that were associated with the layering of the subsurface formations. The geometric attributes were used to recognise the contrasts within the subsurface that were associated with the discontinuities of the subsurface formations. Finally, the inversion attributes were used to recognise the contrasts within the subsurface that were associated with the lithology and porosity of the subsurface. The composition of the predictor space and the function of each of the attributes within that space are presented in Table 1.

Table 1. Composition and functional role of seismic attributes used in the study

Attribute group	Attribute	Notation	Analytical function in reservoir quality prediction
Amplitude	Instantaneous amplitude	A_{inst}	Reflects local reflection strength and contrasts in acoustic properties
	RMS amplitude	A_{RMS}	Characterises the total reflection energy within the analysis window
Frequency	Instantaneous frequency	f_{inst}	Reflects variations in layer thickness and vertical heterogeneity
	Sweetness	SW	Represents the combined amplitude and frequency structure of the signal
Geometric	Coherence	C	Detects discontinuities, body boundaries, and loss of continuity
	Variance	Var	Enhances local heterogeneity and internal segmentation
Inversion	Acoustic impedance	AI	Reflects elastic contrasts associated with lithology and porosity

Note: the predictor space presented in Table 1 corresponds to the final attribute subset retained after pairwise correlation screening and recursive feature elimination

Source: compiled by the authors based on the Project Penobscot dataset, seismic attribute extraction and interpretation performed in OpendTect and Petrel

As shown in Table 1, the constructed predictor space combined four complementary groups of features, none of which duplicated another in functional meaning. The amplitude attributes provided a basic characterisation of reflection strength and proved to be the most informative for identifying acoustic contrasts between intervals. This indicated their particular suitability for delineating zones of lithological change and intervals with differing reservoir quality. The frequency attributes, in turn, reflected the finer internal organisation of sedimentary bodies and proved important for characterising vertical heterogeneity, layering, and textural features of clastic sequences (Shevko *et al.*, 2021).

A different analytical role was demonstrated by the geometric attributes. Unlike the amplitude and frequency attributes, they proved to be the most suitable for identifying lateral boundaries, continuity disruptions, and internal segmentation of reservoir bodies. This is particularly important for the task of reservoir quality prediction, since the quality of a clastic reservoir is determined not only by its average petrophysical properties but also by the degree of spatial connectivity and heterogeneity (Cristea *et al.*, 2023). The inversion-derived attribute, represented by acoustic impedance, showed the clearest relationship with physically interpretable rock properties and therefore became a key link between seismic response and the prediction of continuous parameters, primarily porosity and permeability. The generalisation of the predictor-space structure showed that amplitude and inversion-derived attributes formed the main basis for estimating ϕ and k, whereas frequency and geometric features increased sensitivity to and , that is, to the internal organisation of the reservoir and facies variability. Therefore, the constructed space X proved sufficient for subsequent machine learning-based reservoir quality prediction, as it captured both direct and indirect seismic manifestations of lithology, heterogeneity, and reservoir architecture. This created a substantive basis for the next stage, namely, comparing the performance of machine learning models in predicting reservoir quality parameters. The resulting predictor structure is relevant to clastic reservoir systems of the Balkan type, where lithological variability, structural fragmentation, and internal heterogeneity require a multi-attribute basis for reservoir quality prediction.

Reservoir quality targets and analytical representation of the response variables

The analytical representation of the system response through the continuous target variables in Equation (2) and the categorical facies variable in Equation (3) indicates that reservoir quality in a clastic reservoir cannot be represented by a single parameter. The resulting structure of the target variables revealed the multi-component nature of reservoir quality, in which quantitative petrophysical properties are combined with lithological and facies-related characteristics. As a result, the system response acquired an integrated form, making it possible to consider reservoir quality not in isolation but as a combination of storage capacity, flow potential, the degree of development of sandy bodies, and the internal spatial differentiation of the reservoir. The composition of the target vector and the functional role of its components are presented in Table 2.

Table 2. Target reservoir-quality parameters used in the study

Parameter	Notation	Parameter description	Engineering role in reservoir characterisation
Porosity	ϕ	Fraction of pore space in the rock	Characterises the storage capacity of the reservoir for fluid accumulation
Permeability	k	Ability of the rock to transmit fluids	Determines flow potential and the possible intensity of fluid movement
Sandiness index	SI	Relative proportion of sandy material within the interval	Reflects the degree of development of potentially reservoir-prone bodies in clastics
Facies class	F	Assignment of the interval to a specific facies type	Allows accounting for spatial and lithological heterogeneity of the reservoir

Source: compiled by the authors based on the Project Penobscot dataset, analytical structuring of target variables performed in Python 3.11

As shown in Table 2, the quantitative core of the resulting representation was formed by porosity ϕ and permeability k. These parameters most directly reflected reservoir quality in terms of fluid storage and flow capacity. Their joint presence in the response vector demonstrated that reservoir characterisation requires the simultaneous consideration of both storage and filtration components. In the absence of either parameter, the assessment of reservoir quality would be incomplete, since high porosity does not necessarily imply high permeability, just as satisfactory permeability is not always accompanied by sufficient pore volume. At the same time, the sandiness index SI served as the parameter that linked the petrophysical response to the lithological structure of the reservoir. Its inclusion in the vector Y showed that variations in reservoir quality have not only a physical but also a clearly expressed sedimentological dimension. Unlike ϕ and k, which describe the properties of the rock as a medium for fluid storage and flow, SI reflected the degree of development of sandy bodies as the most prospective reservoir intervals. As a result, this parameter provided the connection between quantitative quality indicators and the lithological expression of the reservoir environment. The facies variable F complemented this system by introducing a spatial and structural level of interpretation. Its presence in the response vector demonstrated that reservoir quality in clastic deposits is determined not only by the magnitude of individual properties but also by the facies organisation of the environment in which these properties are realised. In this sense, F did not duplicate other components of the vector but defined the context of their spatial manifestation. The facies component thus enabled the transition from a purely parametric assessment to a structured representation of the reservoir as a heterogeneous system with internal differentiation. The generalisation of the resulting representation of the target vector showed that ϕ, k, SI and F form a sufficient and complementary set of indicators for reservoir quality characterisation. Porosity and permeability described its storage-flow core, the sandiness index refined the lithological prospectivity of intervals, and the facies class revealed the internal spatial heterogeneity of the system. Therefore, the continuous target vector Y_c represented in Equation (2) was utilised to define the continuous target variable that would be used to compare the different machine learning models for estimating the continuous reservoir parameters. This multi-component representation of the response space is particularly important for Balkan clastic reservoirs, where reservoir quality is controlled simultaneously by storage capacity, flow potential, sand-body development, and facies-related heterogeneity.

Model performance for prediction of continuous reservoir parameters

The comparison of regression results showed that the predictive accuracy of continuous reservoir-quality parameters was uneven both across models and across target variables. The evaluation metrics defined by equations (4)-(7) made it possible to compare the performance of the models for porosity, permeability, and sandiness index within a single analytical framework. The obtained values of R^2, RMSE and MAE indicated that the response of the models depended on the physical nature of the predicted parameter, with higher stability observed for porosity and sandiness index and lower stability for permeability. The comparative regression results are presented in Table 3.

Table 3. Regression performance of machine learning models for predicting porosity, permeability, and sandiness index

Model	ϕ R^2	ϕ RMSE	ϕ MAE	k: R^2	k: RMSE	k: MAE	SI: R^2	SI: RMSE	SI: MAE
Random Forest	0.84	0.028	0.021	0.71	18.6	13.9	0.81	0.061	0.046
XGBoost	0.88	0.024	0.018	0.76	16.9	12.4	0.85	0.054	0.039
Gaussian Process Regression	0.82	0.031	0.023	0.68	20.3	14.8	0.79	0.066	0.049

Note: the metrics are reported for the test set that was divided during the 80:20 split of the training set. Five-fold cross validation was used during the training of the models to ensure generalisation of the models

Source: compiled by the authors based on the Project Penobscot dataset, data preprocessing, model training, and metric calculation performed in Python 3.11, seismic interpretation support in OpendTect and Petrel

As presented in Table 3, XGBoost model achieved the highest overall predictive performance for each of the reservoir properties, based on the best values of each of the three regression metrics. For the prediction of porosity (ϕ), XGBoost achieved the highest value of the coefficient of determination ($R^2 = 0.88$), as well as the lowest values of the error metrics (RMSE = 0.024; MAE = 0.018). Random Forest also achieved relatively high values of each of the three regression metrics for the prediction of porosity, but with higher error values than XGBoost. Lastly, Gaussian Process Regression achieved an acceptable level of accuracy for the prediction of porosity, but with a lower R^2 value and higher error metrics than both the XGBoost and Random Forest models. The same ranking of the three models for the best performance relative to the prediction of permeability (k) was also achieved by each of the models, but with lower accuracy relative to the prediction of porosity. For permeability prediction, XGBoost achieved the best overall performance, with the highest R^2 value (0.76) and the lowest error metrics (RMSE = 16.9; MAE = 12.4). Random Forest models had slightly lower values for each of these metrics for the prediction of permeability, while Gaussian Process Regression achieved the lowest accuracy for permeability prediction. The relatively low accuracy of permeability prediction compared to porosity is likely due to the fact that permeability is a more difficult parameter to predict directly from seismic attributes than porosity.

For the sandiness index (*SI*), all three models achieved relatively high and strong values of each of the three regression metrics, but with XGBoost achieving the highest values. Overall, however, the accuracy of prediction for the sandiness index was higher for Random Forest and Gaussian Process Regression than for permeability. The sandiness index was more easily predicted than permeability, likely due to the fact that the sandiness index has a more stable relation to the attribute space than does permeability. Thus, each of the three models exhibited different levels of accuracy in the prediction of the three reservoir parameters. Porosity was the most easily predicted parameter with the models, followed by the sandiness index, and then permeability. The accuracy of each model in the prediction of each of these parameters indicates that the seismic attribute space is better able to represent storage (porosity and sandiness) than it is to represent flow (permeability). These differences in performance will inform the following stage of this analysis. The obtained ranking of regression performance is especially relevant for Balkan petroleum settings, where permeability prediction is expected to remain more uncertain than porosity and sandiness due to structural complexity and heterogeneous reservoir architecture.

Classification performance for facies prediction

The classification results showed that the facial variable *F* was reproduced by the model with an acceptable level of accuracy sufficient for further use in spatial interpretation of the reservoir. Evaluation by the metrics presented in formulas (8)-(11) made it possible to establish not only the general level of correctness of the classification, but also the degree of balance of the model between the accuracy of class recognition and the completeness of their selection. The obtained results proved that the facies organisation of the reservoir environment is reflected in the attribute space quite stably, although the accuracy of recognition of individual classes remained different. Comparative values of classification metrics are shown in Figure 1.

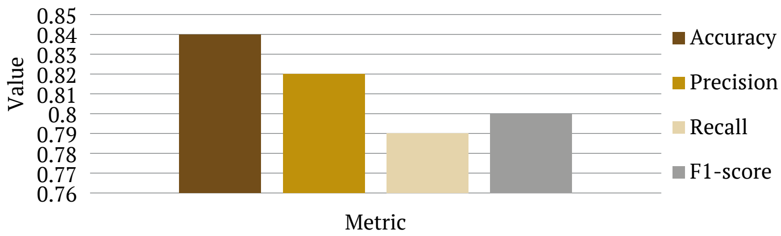

Figure 1. Classification performance metrics for facies prediction

Source: compiled by the authors based on the Project Penobscot dataset, classification performance assessed according to Equations (8)-(11), data processing in Python 3.11

As presented in Figure 1, the classification model achieved an overall accuracy of 0.84 with a precision of 0.82, a recall of 0.79 and an F1-score of 0.8. These results indicate that the model was able to generally stably classify the facies within the reservoir. Furthermore, the relatively balanced precision and F1-score score with the slightly lower recall indicates that the model was generally consistent in its predictions of the facies. Thus, while the model is adequate for predicting the facies within the reservoir, a more detailed analysis is required to determine specifically how the model performed in relation to each individual facies. Overall, the results of classification model indicate that the predictor space is informative enough to be utilised in the prediction of reservoir facies. The accuracy, precision, recall, and F1-score for the model indicate that the model is able to generally return the same structure of the reservoir facies as exist in the analysed reservoir intervals. Moreover, these results allow for the use of the model to predict the reservoir facies as an additional model to those that continuously predict parameters like porosity, permeability, and sandiness index; the determination of reservoir quality is not based upon the individual measurements of each of these parameters, but in relation to the structure of the reservoir facies itself. Furthermore, while these results are generally positive and indicate the ability of the model to appropriately classify the facies within the reservoir, some degree of caution must be applied in interpreting these results; the results are based only upon the classification of the reservoir intervals and does not include a confusion matrix or analyses of the classification of each individual class of reservoir facies. Despite these limitations, the incorporation of a model that predicts reservoir facies is still important in the reservoir description and geological representation of the reservoir. The incorporation of such a model allows for the description and determination of the structure of the analysed reservoir intervals. Finally, the applicability of this specific model and technique to reservoirs of the Balkan type is additionally motivated by the fact that the structure of those reservoirs tends to exhibit the same geological heterogeneity and structure as do the analysed reservoir intervals.

Distribution of predicted porosity, permeability, and sandiness index

After applying the machine learning models to the predictor space, the values of the continuous reservoir-quality parameters, as defined in Equation (2), namely, ϕ, k, and *SI*, were obtained. Unlike the previous subsection that focused upon the various performance metrics of the models, the results of the application of the models to the predictor space will be examined in this subsection. The obtained values of each of these parameters indicates that the predicted reservoir is not uniformly composed of the same type of sedimentary rocks and intervals; rather, there are differences in each of the parameters that indicate the inherent heterogeneity of the reservoir. The statistical characteristics of each of these parameters is summarised in Table 4.

Table 4. Statistical characteristics of predicted reservoir parameters

Parameter	Mean	Standard deviation	Minimum	Maximum	Coefficient of variation
Porosity, ϕ	0.184	0.037	0.112	0.267	0.201
Permeability, k (mD)	146.8	58.4	52.3	284.6	0.398
Sandiness index, SI	0.63	0.14	0.34	0.88	0.222

Source: compiled by the authors based on the Project Penobscot dataset, classification performance assessed according to Equations (7)-(10), data processing in Python 3.11

As shown in Table 4, porosity ϕ proved to be the most homogeneous among the reservoir parameters that were calculated. Porosity has a mean value of 0.184 and a standard deviation of 0.037. These values indicate that storage properties of the reservoir were the least contrastive of the parameters considered. Reservoirs did include intervals with relatively less prospective storage properties than others, and those with relatively better properties, but the contrast between minimum and maximum values was the least among the analysed parameters. This level of homogeneity was also established in the discussion of porosity stability in the previous stage of model evaluation. The homogeneity of this parameter indicates that the reservoir's capacity properties were represented more uniformly throughout the system than its flow-related properties.

Permeability k exhibited a completely different distribution within the reservoir. Permeability had the highest degree of variability of the parameters. Mean permeability value was calculated at 146.8 mD, standard deviation 58.4 mD and the coefficient of variation (a measure of variability relative to the mean) 0.398, the highest of all the calculated coefficients. These values indicate that permeability was the most contrastive parameter within the reservoir space. Minimum permeability of the reservoir was 52.3 mD, and the maximum value was 284.6 mD. Such contrast indicates that permeability is the most sensitive parameter to internal heterogeneity within the reservoir and that its value varies the most among different sections of the reservoir. Sandiness index SI exhibited intermediate variability within the reservoir. Mean sandiness index within the reservoir was 0.63, with values ranging from 0.34 to 0.88. While not as high as permeability, the range in sandiness index indicates differences between the lithologies within the reservoir. The reservoir has segments that exhibit different degrees of sandy bodies. The distributions of porosity, permeability, and sandiness indicate that the reservoir has multi-level heterogeneity within it. Porosity indicates the stability of the reservoir's storage properties, permeability the contrast in the flow properties of the reservoir, and the sandiness the distribution of sandy bodies within the reservoir. These three parameters indicate that the quality of the reservoir within this clastic system is based upon the interrelation of the storage, flow, and sand body properties. This pattern of heterogeneity is especially important to consider within the context of clastic reservoirs of the Balkan type.

Integrated interpretation of reservoir quality and fluid-flow potential

The integrated interpretation of the predicted parameters showed that reservoir quality within the analysed clastic system was controlled by the combined effect of porosity , permeability , sandiness index SI, and facies class F, rather than by any single indicator taken in isolation. When considered together, these variables formed a differentiated reservoir-quality pattern in which storage capacity, transmissibility, lithological prospectivity, and facies-controlled heterogeneity interacted to produce zones with clearly unequal engineering potential. The resulting configuration indicates that the studied reservoir space was internally structured into intervals of high, moderate, and low quality, each associated with a distinct combination of petrophysical and lithological characteristics. The integrated ranking of reservoir zones, obtained from the normalised predicted quality indicators according to Equations (12) and (13), is shown in Figure 2.

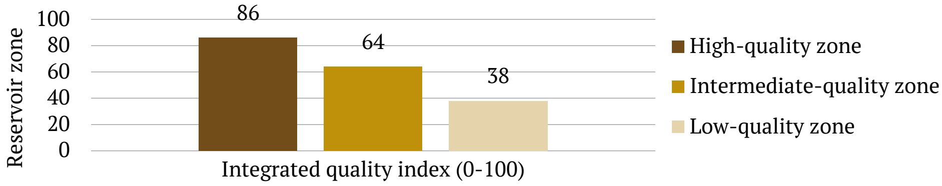

Figure 2. Integrated ranking of reservoir zones by predicted quality indicators
Source: compiled by the authors based on the integrated interpretation of predicted porosity, permeability, sandiness index, and facies class derived from the Project Penobscot dataset, seismic interpretation support in OpendTect and Petrel, data processing in Python 3.11

As shown in Figure 2, the integrated ranking divides the predicted reservoir space into three quality levels. The highest quality of reservoir zones reaches the highest integrated index of 86, the intermediate quality of reservoir zones has an index of 64, while the lowest quality of reservoir zones has the lowest integrated index of 38. The difference between the highest and the lowest ranked zones in terms of the integrated index is 48. The ranking of the quality of reservoir zones within the clastic reservoir system of the Balkan type shows that the prospective zones have higher levels of porosity, permeability, and sand-body development, while the lower ranked zones have lower permeability, porosity, and sand-body development. The integrated ranking of the quality of reservoir zones within the clastic reservoir system of the Balkan type is especially helpful in recognising the effect of the four different variables on the prospectivity of the reservoir zones. Permeability is the most discriminating variable in determining the quality of the reservoir zones, as indicated by the ranking results. The other variables of porosity, sandiness, and facies have other important contributions to the prospectivity of the zone. Porosity is the most important parameter in determining the storage capacity of the reservoir zones, while sandiness and facies are important in recognising the development of sandy bodies within the reservoir zones. The integration of these four variables is important in converting the separate predicted variables of reservoir quality to a model that can help to determine the probable behaviour of the reservoir in relation to the movement of fluids through the reservoir. Furthermore, the developed pattern is especially important in the evaluation of clastic reservoirs of the Balkan type.

The quality of the reservoir zones within these formations are often influenced by lithological variability, variability in the development of sandy bodies within the reservoir zones, and the heterogeneity of the sedimentary facies within the reservoir (Babayeva *et al.*, 2024). Thus, the workflow developed for the recognition of these four variables is relevant to the evaluation of the petroleum reservoirs of Albania and the Balkans region in general, even if the workflow is not applied to the specific data of Albania. Additionally, because of the importance of each of these four variables for the prediction of the quality of the reservoir zones within clastic reservoirs of the Balkan type, the developed workflow is transferable to those same reservoir systems, as well. Thus, the integrated ranking of the quality of reservoir zones within the clastic reservoir system of the Balkan type indicates that the most prospective zones for petroleum exploration and production will have high levels of porosity (ϕ), permeability (k), sand-body development (SI), and favourable facies, while the least prospective zones will contain low levels of those same variables. Thus, the machine learning and seismic attribute workflow developed for prediction of the quality of the reservoir zones within clastic oil fields is analytically suitable for application to those oil fields, including the Balkan-type oil fields, including those within Albania.

Overall, the obtained results demonstrate that the constructed multi-attribute seismic predictor space, combined with machine learning methods, provides a consistent and analytically reliable basis for reservoir quality prediction in clastic oil systems. The results confirmed that reservoir quality cannot be described by a single parameter, but emerges from the interaction of porosity, permeability, sandiness index, and facies structure. The models showed the highest predictive stability for the porosity and sandiness index, while permeability proved to be the most sensitive to heterogeneity within the reservoir rock. The classification of facies within the area also helped to reveal information about the reservoir that was not otherwise obtained from the prediction models. The distribution of the predicted reservoir parameters helped to reveal the different levels of heterogeneity existing within the reservoir, as well as the differing stability of reservoir properties like storage versus flow potential. The differentiated zones within the reservoir have different engineering potentials due to the reservoir characteristics like permeability and sandy bodies. Such an analysis helps to confirm the applicability of this workflow to these types of reservoirs and its relevance to the Balkans-type reservoirs specifically.

Discussion

Comparison of the obtained results with modern works in the field of seismic interpretation made it possible to more precisely outline their place in the current direction of reservoir characterisation, where the combination of multi-attribute analysis with machine learning algorithms is gaining more and more importance. Such a comparison revealed not only the commonality of methodological approaches, but also differences in which parameters of the reservoir were reproduced most stably, how important the role of facies control remained, and how the influence of the internal heterogeneity of the environment was interpreted. In this context, it is advisable to consider the obtained results sequentially, in the order of the involved sources, comparing the nature of the established regularities with the conclusions of the conducted research.

An increase in the accuracy of estimation of petrophysical properties in a lithologically variable environment was obtained in the work of J.O. Olutoki *et al.* (2024) by combining geostatistical inversion with extreme gradient boosting. A similar pattern was found in the conducted study, where XGBoost provided the best results for predicting porosity and sandiness index, which indicated the high efficiency of this algorithm in the tasks of detecting non-linear relationships between seismic attributes and reservoir properties. An increase in interpretive reliability under conditions of limited direct control was shown in a study by N. Abd-Elfattah *et al.* (2025), where machine learning algorithms were combined with the amplitude versus offset class II workflow for hydrocarbon prospectivity assessment. The obtained results revealed a close logic, since the multicomponent space of seismic predictors also provided stable reproduction of reservoir characteristics even in the presence of a complex internal structure and heterogeneous distribution of properties within the clastic system. Improvement of identification of promising intervals under conditions of structural and stratigraphic complexity was achieved in the work of M. Reda *et al*. (2024) by combining high-resolution seismic analysis and 3D modelling. In the conducted research, a related effect was obtained at the stage of integrated ranking of zones, when the greatest informativeness was provided not by the isolated interpretation of individual attributes, but by the combined analysis of porosity, permeability, sandiness index and facies structure. A more meaningful facies interpretation under the conditions of integration of mineralogical and porosity information was demonstrated in the work of P. Owusu *et al*. (2025), where machine learning was used for seismofacies analysis. Such an approach was consistent with the obtained results, as the facies classification did not duplicate the prediction of continuous parameters, but extended it by adding spatial context and clarifying the internal organisation of the reservoir.

A reduction in predictor redundancy and an improvement in the identification of sand bodies was established in a study by T.F. Ren *et al*. (2025), where Light Gradient Boosting Machine was combined with RFECV. The same methodical advantage was observed in the conducted study, since preliminary correlation screening and recursive feature elimination made it possible to form a compact and functionally disjoint set of attributes, on which the sandiness index was predicted more stable than permeability. Increased sensitivity to fine sand bodies in complex sedimentation conditions was obtained by J. Wang *et al*. (2025) thanks to the combination of self-organising map-optimisation of attribute space with seismic waveform inversion. The obtained results revealed a close trend, since it was the combination of amplitude, frequency, geometric and inversion indicators that provided acceptable sensitivity to lithological variability and internal heterogeneity of clastic intervals. The need for an integrated approach to reservoir characterisation was substantiated by G. Murtaza *et al*. (2025), who combined petrophysical, seismic and rock-physics analysis and showed the limitation of relying on only one type of data. The obtained results developed the same logic, since the quality of the reservoir was interpreted not through a single parameter, but through the interaction of porosity, permeability, sandiness and facies belonging within a single analytical scheme. The dependence of the reliability of seismofacies on the consideration of the wider sedimentary context was shown in the work of P.C. Kumar *et al*. (2026), where subsurface and outcrop data were integrated for the analysis of complex sedimentation systems.

A similar trend was found in the conducted study, where the interpretive value of the predicted parameters increased when they were considered not in isolation, but within the integral internal architecture of the clastic system. An improvement in the prediction of lithofacies and their spatial distribution in a deltaic reservoir was demonstrated by S.M. Badr *et al*. (2026) by integrating machine learning with seismic attributes. This was consistent with the obtained results, since it was the facies classification that made it possible to move from a set of individual predicted values to a spatially differentiated model of the reservoir with different degrees of perspective. An increase in the geological relevance of the prediction of reservoir properties was obtained in the work of K. Ren *et al*. (2026) by optimising seismic attributes in combination with spectral decomposition and automated machine learning. A similar conclusion emerged from the conducted study, where the best prognostic quality was achieved not due to the maximum expansion of the descriptor space, but due to the selection of features that reflected different, but not duplicated aspects of lithology, heterogeneity and reservoir architecture. A significant improvement in the prediction of sand bodies in complex structural and sedimentary conditions was shown by J. Liu *et al*. (2026), who combined multi-attribute seismic analysis with machine learning under facies constraint. The obtained results revealed the same pattern, as it was the inclusion of the facies component together with porosity, permeability and sandiness index that made it possible to move from individual predictions to a holistic model of reservoir quality with a clear separation of high, medium and low potential zones.

Improved accuracy of porosity estimation in heterogeneous siliciclastic rocks was obtained by M. Hussain *et al*. (2025) through the integration of conventional approaches, machine learning, and seismic inversion. A similar tendency was observed in the conducted study, where porosity proved to be the parameter with the highest predictive stability, while the combination of multi-attribute analysis and machine learning algorithms produced the best accuracy values specifically for this indicator. High efficiency of machine learning for predicting lithological and petrophysical parameters in hydrocarbon exploration tasks was demonstrated by D. Arkadiusz *et al*. (2025) using the Carpathian Foredeep case study. The obtained results revealed a similar pattern, as they also indicated that machine learning was most effective when applied not to the isolated prediction of a single variable, but to the simultaneous interpretation of several interrelated reservoir characteristics. The integration of seismic attributes and Hjorth parameters for seismic facies classification and total organic carbon prediction was implemented by R. Al Hasan *et al*. (2025), who showed that the expansion of the descriptor space increased interpretive sensitivity to the internal heterogeneity of the medium. A comparable effect was obtained in the conducted study, where a heterogeneous set of amplitude, frequency, geometric, and inversion-derived attributes made it possible to reproduce both continuous reservoir parameters and facies structure within a single analytical framework. The issue of training-data consistency and the need for its verification in the context of seismic deep learning tasks was examined in detail by A. Pradhan & T. Mukerji (2022). The authors showed that the quality of the predictions of the deep learning models depended on two factors: the validity of the training data and the absence of contradictions between the training data and geological expectations. These findings correspond to the study conducted herein, as the same analytical sample of seismic data was formed only from data segments with consistent time-depth relationships and signal quality, as well as good seismic well log ties.

A new approach to seismic-attribute optimisation based on forward modelling and machine learning was proposed by W. Li *et al*. (2023), who found that targeted feature selection improved the prediction quality of fluvial reservoirs. A similar conclusion followed from the obtained results, since the best predictive performance was achieved not through mechanical expansion of the attribute space, but through the selection of predictors representing different aspects of lithology, heterogeneity, and reservoir architecture without functional duplication. Petroleum prospect generation and evaluation based on machine learning and seismic attributes were examined by M. Farfour *et al*. (2025), who showed that data-driven analysis improved the quality of prospect evaluation in offshore conditions. The results obtained made it possible to not only rank the determined zones according to their quality but also to create a more comprehensive model of the reservoir based on these parameters.

A similar approach to the identification of prospective reservoirs and the risking of their exploration was presented by M. Farfour & D. Foster (2022). The authors used machine learning algorithms to predict a series of seismic attributes that made it possible to better explore the area of interest. The same principle can be seen in the development of a model that predicted lithology that provided more information than each of the individual seismic attributes alone. The use of machine learning to predict lithology based on well log data was performed by W. Jia *et al*. (2026). In their study, the authors found that the use of machine learning made it possible to stably reconstruct the lithological features of the explored area based on the inversion of the seismic data used to train the model. As such, it is not surprising that the inverted component of the model for reservoir quality was also essential to determine the lithology of the reservoir, mainly its porosity.

The capabilities of unsupervised machine learning combined with multi-attribute analysis for identifying low-saturation gas reservoirs were demonstrated by J. Chenin & H. Bedle (2022), who proved that even without rigidly predefined training labels, a multidimensional attribute space could reveal geologically meaningful patterns. A comparable observation was obtained in the conducted study, since facies differentiation and spatial zoning of reservoir quality also confirmed that the multi-attribute space contained sufficient information to reproduce the internal organisation of the clastic system.

The generalisation of the comparison with these sources showed that the obtained results extended the contemporary direction of integrated seismic interpretation oriented toward the combination of inversion, attribute optimisation, training-data quality control, and machine learning algorithms. At the same time, the conducted study specified that, for clastic systems, the most stable predictions were obtained for parameters more closely related to storage properties and lithological structure, whereas permeability remained the most sensitive to internal heterogeneity, segmentation, and the complexity of reservoir architecture.

Conclusions

The study showed that machine learning methods provide a viable and analytically sound basis for reservoir quality prediction in clastic oil fields. Reservoir quality could not be reliably characterised by any single seismic attribute, since it was controlled by the combined influence of storage capacity, flow potential, sand-body development, and facies characteristics. At the same time, the multi-attribute predictor space proved adequate for representing the main parameters of reservoir quality. The efficacy of various models was determined to be dependent upon the physical nature of the parameter that was to be predicted. The stability of the reservoir parameters was highest for the prediction of porosity, with the XGBoost algorithm achieving the highest R^2 value of 0.88, the lowest RMSE of 024, and the lowest MAE of 0.018. The sandiness index followed closely behind with an R^2 of 0.85, an RMSE of 0.054, and an MAE of 0.039. The permeability of the reservoir was the hardest to predict, with the best performing model achieving only an R^2 of 0.76, an RMSE of 16.9, and an MAE of 12.4. Finally, the classification of the reservoir facies exhibited a relatively steady quality in its predictions, with accuracies of 0.84, precision of 0.82, recall of 0.79, and an F1 score of 0.80.

Thus, the proposed methodology was applicable both to continuous prediction of reservoir properties and to facies classification. The predicted parameter distribution confirmed reservoir heterogeneity: porosity was the most stable parameter, whereas permeability showed the highest variability. The integrated assessment identified high-, moderate-, and low-quality intervals with indices of 86, 64, and 38, respectively. The most promising intervals were associated with higher porosity, higher permeability, stronger sand-body development, and more favourable facies characteristics. Future research in this area would be directed towards testing the accuracy of this methodology upon a dataset of reservoirs within Albania, collecting additional attributes related to seismic reflection and inversion analyses, increasing the accuracy in the prediction of permeability within those reservoirs, and performing uncertainty analyses regarding the quality predictions of those reservoirs.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

[1] Abd-Elfattah, N., Dahroug, A., El Kammar, M., & Fahmy, R. (2025). Machine learning and AVO class II workflow for hydrocarbon prospectivity in the Messinian offshore Nile Delta Egypt. *Scientific Reports*, 15(1), article number 3566. doi: 10.1038/s41598-025-86765-7.

[2] Al Hasan, R., Saberi, M.H., & Riahi, M.A. (2025). Seismic facies classification and TOC prediction using seismic attributes and Hjorth parameters, in the Groningen field in northeastern Netherlands. *Discover Applied Sciences*, 7(10), article number 1202. doi: 10.1007/s42452-025-07794-5.

[3] Alexandridis, I., Oikonomopoulos, I.K., Carvajal-Ortiz, H., Gentzis, T., Kalaitzidis, S., Georgakopoulos, A.N., & Christanis, K. (2022). Discovery of a new source-rock interval within the Pantokrator Formation, Ionian Zone, western Greece: Insights from sulfur speciation and kinetics analyses. *Marine and Petroleum Geology*, 145, article number 105918. doi: 10.1016/j.marpetgeo.2022.105918.

[4] Ali, M., Changxingyue, H., Wei, N., Jiang, R., Zhu, P., Hao, Z., Hussain, W., & Ashraf, U. (2024). Optimizing seismic-based reservoir property prediction: A synthetic data-driven approach using convolutional neural networks and transfer learning with real data integration. *Artificial Intelligence Review*, 58(1), article number 31. doi: 10.1007/s10462-024-11030-8.

[5] Arkadiusz, D., Tomasz, T., Anita, L.Ś., & Krzysztof, S. (2025). Machine learning approaches for predicting lithological and petrophysical parameters in hydrocarbon exploration: A case study from the Carpathian foredeep. *Energies*, 18(17), article number 4521. doi: 10.3390/en18174521.

[6] Babayeva, T., Guliyev, A., İslamzade, T., İslamzade, R., Haciyeva, X., Ashurova, N., Aliyeva, A., & Maksudov, S. (2024). Impacts of irrigation with Cd-contaminated water from Sugovushan Reservoir, Azerbaijan on total cadmium and its fractions in soils with varied textures. *Eurasian Journal of Soil Science*, 13(2), 145-152. doi: 10.18393/ejss.1424421.

[7] Badr, S.M., Mohamed, F.H., Mansour, A.S., & Elessawy, A.A. (2026). Machine learning integration with seismic attributes for lithofacies prediction and distribution in Abu Madi reservoir, onshore Faraskour gas field, east Nile Delta, Egypt. *Journal of Petroleum Exploration and Production Technology*, 16(2), article number 24. doi: 10.1007/s13202-025-02125-1.

[8] Chen, Y., Zhao, L., Pan, J., Li, C., Xu, M., Li, K., Zhang, F., & Geng, J. (2021). Deep carbonate reservoir characterisation using multi-seismic attributes via machine learning with physical constraints. *Journal of Geophysics and Engineering*, 18(5), 761-775. doi: 10.1093/jge/gxab049.

[9] Chenin, J., & Bedle, H. (2022). Unsupervised machine learning, multi-attribute analysis for identifying low saturation gas reservoirs within the deepwater Gulf of Mexico, and Offshore Australia. *Geosciences*, 12(3), article number 132. doi: 10.3390/geosciences12030132.

[10] Cristea, V.-M., Baigulbayeva, M., Ongarbayev, Y., Smailov, N., Akkazin, Y., & Ubaidulayeva, N. (2023). Prediction of oil sorption capacity on carbonized mixtures of Shungite using artificial neural networks. *Processes*, 11(2), article number 518. doi: 10.3390/pr11020518.

[11] El-Dabaa, S.A., Metwalli, F.I., Maher, A., & Ismail, A. (2024). Unsupervised machine learning-based multi-attributes analysis for enhancing gas channel detection and facies classification in the serpent field, offshore Nile Delta, Egypt. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 10(1), article number 185. doi: 10.1007/s40948-024-00907-1.

[12] Farfour, M., & Foster, D. (2022). Machine learning and seismic attributes for prospect identification and risking: An example from offshore Australia. In *SEG international exposition and annual meeting* (article number D011S011R004). Houston: OnePetro. doi: 10.1190/image2022-3752064.1.

[13] Farfour, M., Hedjam, R., Foster, D., & Gaci, S. (2025). Machine learning and seismic attributes for petroleum prospect generation and evaluation: An example from offshore Australia. *Interpretation*, 13(4), 805-815. doi: 10.1190/INT-2024-0177.1.

[14] Ferreira, D.J., Dias, R.M., & Lupinacci, W.M. (2021). Seismic pattern classification integrated with permeability-porosity evaluation for reservoir characterization of presalt carbonates in the Buzios Field, Brazil. *Journal of Petroleum Science and Engineering*, 201, article number 108441. doi: 10.1016/j.petrol.2021.108441.

[15] Fociro, A., Fociro, O., Prifti, I., Muçi, R., & Pekmezi, J. (2023). Petrography and mineralogy of the Oligocene flysch in Ionian Zone, Albania: Implications for the evolution of sediment provenance and paleoenvironment. *Open Geosciences*, 15(1), article number 20220442. doi: 10.1515/geo-2022-0442.

[16] Hussain, M., Shakir, U., Naseer, Z., Amjad, M.R., Arshad, F., Kamal, M., & Radwan, A.E. (2025). Integrating conventional and machine learning aproaches with seismic inversion for optimized porosity estimation in heterogeneous siliciclastic rocks. *Physics and Chemistry of the Earth, Parts A/B/C*, 142, article number 104248. doi: 10.1016/j.pce.2025.104248.

[17] Jia, W., Liu, D., Zhang, G., Lan, T., Sun, Z., & Hu, H. (2026). Lithology prediction based on well-log data and seismic inversion result: A machine learning approach. *Journal of Applied Geophysics*, 247, article number 106131. doi: 10.1016/j.jappgeo.2026.106131.

[18] Kumar, P.C., Bedle, H., Kumar, J., Goswami, T.K., & Sain, K. (2026). Seismic facies characterization: Integrated subsurface-outcrop analysis for complex depositional systems in northeast India. *Artificial Intelligence in Geosciences*, 7(1), article number 100196. doi: 10.1016/j.aiig.2026.100196.

[19] Li, W., Yue, D., Colombera, L., Duan, D., Long, T., Wu, S., & Liu, Y. (2023). A novel method for seismic-attribute optimization driven by forward modeling and machine learning in prediction of fluvial reservoirs. *Geoenergy Science and Engineering*, 227, article number 211952. doi: 10.1016/j.geoen.2023.211952.

[20] Liu, J., Lin, C., Elders, C., & Faris, A. (2026). Sandbody prediction based on fusion of seismic multi-attributes and machine learning under sedimentary facies constraint – a case study of Chenguanzhuang area in Dongying depression, Bohai Bay basin. *Applied Sciences*, 16(7), article number 3341. doi: 10.3390/app16073341.

[21] Murtaza, G., Ali, N., Hussain, W., Iqbal, S.M., Azhar, M.U., Ahmad, T., Iltaf, K., & Ahmed, A. (2025). Integrating petrophysical, seismic and rock physics analyses for precise reservoir characterization. *Earth Systems and Environment*, 9(3), 2165-2187. doi: 10.1007/s41748-025-00654-7.

[22] Narayan, S., Sahoo, S.D., Kar, S., Pal, S.K., & Kangsabanik, S. (2024). Improved reservoir characterization by means of supervised machine learning and model-based seismic impedance inversion in the Penobscot field, Scotian Basin. *Energy Geoscience*, 5(2), article number 100180. doi: 10.1016/j.engeos.2023.100180.

[23] Oitseva, T. (2024). Geological structure and mineralogical composition of the Kara-Ayak rare metal ore occurrence (East Kazakhstan). In *International multidisciplinary scientific geoconference surveying geology and mining ecology management SGEM 2024* (pp. 71-78). Vienna: SGEM. doi: 10.5593/sgem2024/1.1/s01.10.

[24] Olutoki, J.O., Elsaadany, M., Siddiqui, N.A., Haque, A.E., Ali, S.H., Rashid, A., & Akinyemi, O.D. (2024). Estimating petrophysical properties using geostatistical inversion and data-driven extreme gradient boosting: A case study of late eocene McKee formation, Taranaki Basin, new Zealand. *Results in Engineering*, 24, article number 103494. doi: 10.1016/j.rineng.2024.103494.

[25] Owusu, P., Raef, A., & Sharaf, E. (2025). Carbonate seismic facies analysis in reservoir characterization: A machine learning approach with integration of reservoir mineralogy and porosity. *Geosciences*, 15(7), article number 257. doi: 10.3390/geosciences15070257.

[26] Pradhan, A., & Mukerji, T. (2022). Consistency and prior falsification of training data in seismic deep learning: Application to offshore deltaic reservoir characterization. *Geophysics*, 87(3), 45-61. doi: 10.1190/geo2021-0568.1.

[27] Rahma Putra, M.H., Hermana, M., Yogi, I.B., Hossain, T.M., Abdurrachman, M.F., & Kadir, S.J. (2024). Reservoir porosity assessment and anomaly identification from seismic attributes using Gaussian process machine learning. *Earth Science Informatics*, 17(2), 1315-1327. doi: 10.1007/s12145-024-01240-7.

[28] Reda, M., El-Gendy, N.H., Raef, A., Elmashaly, M.M., AlArifi, N., & Barakat, M.K. (2024). Advancing neogene-quaternary reservoir characterization in offshore Nile Delta, Egypt: high-resolution seismic insights and 3D modeling for new prospect identification. *Environmental Earth Sciences*, 83(20), article number 581. doi: 10.1007/s12665-024-11846-1.

[29] Ren, K., Li, W., Zhao, L., Wang, W., Wang, J., Wang, H., Li, Y., Qu, L., & Yue, D. (2026). Seismic-attribute optimization for improving reservoir prediction integrating the spectral decomposition and automated machine learning. *Journal of Applied Geophysics*, 248, article number 106167. doi: 10.1016/j.jappgeo.2026.106167.

[30] Ren, T.F., Feng, Z.B., Zhang, Y., Zhang, X., Jiang, L., Ning, Y.L., Wang, J., Ding, J., & Qi, Z.S. (2025). A seismic multi-attribute sandbody identification method based on the LightGBM-RFECV coupling algorithm. *Applied Geophysics*, 22(3), 757-769. doi: 10.1007/s11770-025-1292-y.

[31] Riyadi, Z.A., Olutoki, J.O., Hermana, M., Latif, A.H., Yogi, I.B., & Kadir, S.J. (2024). Machine learning prediction of permeability distribution in the X field Malay Basin using elastic properties. *Results in Engineering*, 24, article number 103421. doi: 10.1016/j.rineng.2024.103421.

[32] Shevko, V.M., Zharmenov, A.A., Aitkulov, D.K., & Terlikbaeva, A.Z. (2021). Complex processing of oxidized copper and zinc oxide ores with simultaneous production of several products. *Physicochemical Problems of Mineral Processing*, 57(1), 226-249. doi: 10.37190/ppmp/131091.

[33] Soto, J.I., *et al*. (2024). Contrasting styles of salt-tectonic processes in the Ionian Zone (Greece and Albania): Integrating surface geology, subsurface data, and experimental models. *Tectonics*, 43(1), article number e2023TC008104. doi: 10.1029/2023TC008104.

[34] Wang, J., Guan, D., Huang, X., He, Y., Li, H., Xu, W., Liu, R., & Feng, B. (2025). Integrated SOM multi-attribute optimization and seismic waveform inversion for thin sand body characterization: A case study of the paleogene lower E3d2 sub-member in the HHK depression, Bohai Bay basin. *Applied Sciences*, 15(9), article number 5134. doi: 10.3390/app15095134.

Аналіз сейсмічних атрибутів з використанням методів машинного навчання для прогнозування якості колекторів на родовищах осадових порід

Фахрі Махо

Доктор філософії у галузі водної інженерії, викладач
Політехнічний університет Тирани
1001, вул. Теодора Кеко, 28, м. Тирана, Албанія
https://orcid.org/0009-0002-0673-4577

Енкелейда Джиналі

Доктор філософії у галузі водної екологічної інженерії, доцент
Політехнічний університет Тирани
1001, вул. Теодора Кеко, 28, м. Тирана, Албанія
https://orcid.org/0009-0005-4942-3893

Анотація. Метою дослідження було оцінити ефективність аналізу сейсмічних атрибутів з використанням методів машинного навчання для прогнозування якості колекторів на кластичних та балканських нафтових родовищах. Валідаційне дослідження було проведено на основі загальнодоступних сейсмічних даних та даних про колектори в рамках проекту «Penobscot» для низки атрибутів у просторі предикторів, а також для чотирьох змінних якості колекторів. XGBoost показав найкращі результати регресії для всіх безперервних параметрів, з найвищою точністю для пористості (R^2 = 0,88, середньоквадратична похибка = 0,024, середня абсолютна похибка = 0,018), за якою слідував індекс піщаності (R^2 = 0,85, середньоквадратична похибка = 0,054, середня абсолютна похибка = 0,039), тоді як проникність залишилася найскладнішим для прогнозування параметром (R2 = 0,76, середньоквадратична похибка = 16,9, середня абсолютна похибка = 12,4). Класифікація фацій також продемонструвала стабільні результати: точність 0,84, прецизія 0,82, відтворюваність 0,79 та F1-показник 0,8. Розподіл прогнозованих параметрів виявив неоднорідну систему колектора, де пористість була найстабільнішим параметром, індекс піщаності демонстрував середню мінливість, а проникність – найвищий контраст. Комплексна інтерпретація прогнозованих показників дозволила розділити пласт на зони високої, середньої та низької якості з індексами 86, 64 та 38 відповідно. Отже, якість пласта в першу чергу визначається факторами, пов'язаними з потенціалом накопичення та протікання, піщаними тілами та фаціями, проте методи, доповнені технологіями машинного навчання, можуть ефективно допомогти інженерам-нафтовикам у складанні точних прогнозів щодо пластів у цих неоднорідних системах. Практична цінність дослідження полягає в можливості використання розробленого алгоритму для ефективного пошуку та визначення пріоритетності ділянок видобутку нафти й газу в родовищах за допомогою машинного навчання, зокрема в геологічно та структурно складних осадових системах

Ключові слова: пористість; середньоквадратична похибка; проникність; потенціал протікання; індекс піщаності; балканський тип