ANAS Transactions **Earth Sciences** **1 / 2025**

http://www.journalesgia.com

APPLICATION OF MACHINE LEARNING METHODS AND NEURAL NETWORKS IN THE OBJECTIVES OF LITHOFACIES MAPPING AND RESERVOIR PROPERTIES ASSESSMENT: ANALYSIS AND SELECTION OF METHODS

Abetov A.E.[1], Seitzhanov A.K.[2], Samenov Ye.R.[1]

[1]*Satbayev University, Kazakhstan*
22, Satpayev Str., Almaty, 050013
[2]*Kazakh-British Technical University, Kazakhstan*
59, Tole Bi Str., Almaty, 050000: ansar.seitzhanov.98@mail.ru

Keywords: *artificial intelligence, machine learning, classification of geological problems, linear and polynomial regression, clusterization of lithofacies, reservoir properties*

Summary. The paper discusses the application of artificial intelligence (AI) methods to address challenges in lithofacies mapping and the assessment of reservoir properties. The choice of an AI method depends on the nature of the data, objectives of the study (such as classification, regression, clustering, or image segmentation), and requirements on the final interpretation and modeling results. An analysis of various machine learning (ML) algorithms including the support vector machine (SVM), random forest (RF), neural networks, etc. were conducted. Evaluated effectiveness of each method was evaluated on the basis of open-source data and geological datasets. Advantages and disadvantages of these methods were analyzed and factors influencing on the selection of an appropriate AI method were identified. Classification of geological problems and corresponding AI methods, encompassing SVM, RF, linear and polynomial regression, k-means clustering, hierarchical clustering, and convolutional neural networks (CNN) were presented. The article also introduces open-source ML platforms such as TensorFlow, PyTorch, and Keras along with factors influencing on the selection of the optimal AI method for lithofacies analysis and reservoir property assessment. Recommendations to select the most suitable AI methods for specific objectives were provided. The importance of data volume and quality in selection of AI method and prevention of model overfitting was emphasized.

1. Introduction

In modern petroleum geology, where data volume increases exponentially, artificial intelligence (AI) becomes an essential tool to solve complex problems. Traditional methods based on the interpretation and modeling of field and well geophysical data, as well as laboratory studies of core, cuttings, and fluid properties are usually labor-intensive, time-consuming, subjective, and limited in their ability to process massive data.

The labor intensity and time consumption are determined by: a) the level of professional expertise of specialists, the availability of equipment, methodological complexes, and software; b) insufficient flexibility of workflows; c) the necessity to write additional scripts for software to optimize the process.

Subjective limitations are determined by: a) overestimated expectations from clients and/or insufficient competence of the performers; b) the need to configure machine learning (ML) based on the results of "expert" training. The system will not create something entirely new; it will attempt to replicate the expert's solution. This is relevant while considering the specifics of commercial software.

Limited capabilities to process big data sets addressed as positive ML use cases emerge. The advanced software already incorporates «Python» as one of the programming modules/languages, which has significantly expanded their capabilities.

According to the Technological Research of KPMG (2025), AI in Central Asia and the Caucasus is not yet a priority technology for most companies. Only 17% of organizations actively use AI, while 29% face insufficient management support and investment constraints, which hinders its implementation.

Moreover, only 44% of companies in Central Asia and the Caucasus plan to implement AI for automating routine targets and improving employee productivity. This indicates a growing interest in using AI as a tool for optimizing operational processes and reducing costs. Additionally, 36% of organizations aim to apply AI for developing new products and services highlighting its role in maintaining competitive advantage and stimulating innovation (KPMG, Caucasus and Central Asia, 2025).

In this context, increasing attention was paid to the application of AI methods, which have the ability to rapidly process big data and identify patterns opening new opportunities for more accurate, efficient, and objective forecasting of oil and gas potential in exploration targets. Of particular relevance is the use of AI to solve objectives related to lithofacies mapping and the assessment of reservoir properties, which play a key role in success of geological exploration in any oil and gas prospective region.

Therefore, the goal of this study was to analyze and determine the capabilities of AI methods for solving lithofacies mapping and reservoir property assessment targets, with a focus on their practical application based on a comprehensive analysis of available archival and published data and requirements for the results.

An important role in this process was played by the analysis and selection of optimal AI methods depending on specific targets and data characteristics. Significant attention was given to the correctness of the initial data, the algorithms of their preparation, and the configuration of models. This is particularly relevant since not all objectives have been fully resolved, especially when it comes to deep convolutional networks. To overcome the identified challenges, it recommended to empirically assembling such networks to address the following objectives:

- Review of existing AI methods applied in petroleum geology for lithofacies mapping and the assessment of reservoir properties.
- Classification of AI methods based on the types of objectives to solve and their functionality.
- Determination of criteria to select the optimal AI method depending on the nature of the data, research objectives, available resources, and requirements for the interpretability of results.
- Analysis of examples of successful applications of AI methods in oil and gas industry.

2. Research methods and data

The modern world of AI in petroleum geology characterized by openness and active exchange of ideas. There are numerous open-source ML platforms to create AI systems, such as TensorFlow, PyTorch, Keras, MxNet, CNTK, Caffe, Paddle, Scikit-learn, and Weka, which significantly simplify and accelerate the process of interpretation and modeling geological, geophysical, and field data (Ng, 2023).

For example, Keras is a software interface or library that allows building deep networks and simplifies working with the core functions of TensorFlow making them more user-friendly and accessible. It provides high-level abstractions, automates routine processes, and reduces the complexity of writing code while retaining the full power and flexibility of the original framework (Keras.io/).

Scikit-learn is a library for data processing. It can solve most objectives in conjunction with Pandas. Pandas is a Python software library for data processing and analysis (Scikit-learn official documentation).

In addition to ML, there are nonlinear neural networks that enable inverse-predictive modeling during the integrated interpretation of seismic, drilling, and other data. In such cases, big data allow the use of more complex algorithms, which require a significant number of examples of training.

In these scenarios, it is crucial to consider the issue of model overfitting, a common problem in ML where the model "memorizes" the training data too well, including noise and random deviations, instead of identifying general patterns in the real geological environment. As a result, such a model performs excellently on the training dataset but poorly on new data (Easyoffer.ru/question/5440).

Combating overfitting is a classic step in tuning ML algorithms. Through a Pipeline (an automated data processing workflow that includes collection, cleaning, transformation, model training, and result evaluation), the data samples obtained after the analytical stage are divided into tuning and blind test sets. The tuning dataset were divided for internal cross-validation. This helps prevent overfitting, as only the parameters that yield the best results according to the search metric are selected during the model tuning phase~~,~~ discarding the overfitting stage.

The problem of model overfitting due to insufficient data (Easyoffer.ru/question/5440) can also be addressed by applying regularization methods, such as ridge regression. Regularization introduces a penalty term into the loss function of the model reducing its tendency to overfit. Essentially, ridge regression implements L2-regularization. This penalty term limits the magnitude of the model's parameters by minimizing their squared values, thereby reducing model complexity and enhancing its generalization ability (Hastie et al., 2009). This approach becomes particularly important in situations with limited data availability, which is typical for many studies where the existing information is often insufficient to build reliable models

In addition to traditional neural networks, a new type of neural network based on the Kolmogorov-Arnold mathematical theorem (Kolmogorov-Arnold network – KAN) for representing multivariate functions gains popularity. According to this theorem, any continuous function of several variables can represent as a superposition (composition) of continuous functions of one variable followed by an addition operation (Simplified explanation of the new Kolmogorov-Arnold network from MIT, 2024).

Due to the improved approximation and interpretability, KAN can be used for: a) building models that classify lithofacies and reservoir properties based on well logging and seismic data; b) identifying complex relationships between geophysical parameters and lithofacies types.

At the same time, the choice of the optimal AI method or methods for lithofacies analysis and the assessment of reservoir properties determined by a combination of factors related to the nature of the data, research objectives, available resources, and interpretation requirements.

Data presented as numerical arrays (e.g., well logging values, results of core and thin section laboratory analyses, seismic data, electrical, gravity, and magnetic survey data) play a key role in selection of the AI method.

The methodology of the conducted research involves the analysis and selection of AI methods based on the following factors:

- Nature of the data (type, volume, quality).
- Research objectives (classification, regression, clustering, image segmentation).
- Requirements for the interpretability of results.
- Available resources.

This article provides an overview of modern methods to apply AI in lithofacies analysis and the assessment of reservoir properties. Various approaches based on ML, neural networks, and other AI algorithms are discussed.

3. Research results

In the practical assessment of lithofacies and reservoir properties, data from seismic surveys, well logging, and laboratory core studies (petrophysical properties, lithological descriptions), as well as geological data on the study area~~,~~ are used.

Next, the presented overview of ML methods was applied to solve various geological targets. The methods were classified by their functionality with a brief description of their essence and geological examples.

Various ML algorithms can be used to solve the objectives, such as:

- **Support Vector Machines (SVM):** Effective for classification and regression in complex environments. Classification was performed: a) based on anomalies on geophysical maps; b) at separating rock samples into different types based on geochemical, petrographic, or geophysical data; c) determining the probability of hydrocarbon presence based on geological, geophysical, and field data.
- **Random Forest (RF)**: An ensemble of decision trees that provides high accuracy and resistance to overfitting by using random subsets of features and training data for each case.
- **Neural Networks**: Capable of approximating complex nonlinear relationships using inversion techniques to calculate the contrast volume of geophysical parameters around the target layer improving the correlation of target parameters.

The prediction procedure consists of two stages: 1) Training: The stage of training neural networks using paired input data (attributes of potential geophysical fields and sets of map values in sliding windows) with the calculation of optimal neural network coefficients by minimizing the objective function; 2) Computation: The stage of calculating the attribute of potential geophysical fields based on a set of the given maps.

The output includes maps of standard deviation P10, P50, and P90 (pessimistic, realistic, and optimistic forecasts). This approach considered valid when dealing with a discrete parameter, such as reservoir/non-reservoir. In this case, the model outputs a probability. Subsequently, the probability distribution can be converted into an array indicating the presence or absence of a reservoir, based on a defined threshold value.

However, if the parameter is continuous, the solution search will rely on the standard approach of minimizing the residual functional of multiple variables.

Artificial Neural Network (ANN): Consists of multiple interconnected "neurons" organized in layers. It is capable of approximating complex nonlinear relationships between input and output data.

- Convolutional Neural Networks (CNN): Well-suited for processing seismic data to identify structures and classify rock types.
- Recurrent Neural Networks (RNN): Used for processing sequential data to identify patterns, such as modeling changes in mineral concentrations.
- Generative Adversarial Networks (GAN): Used for generating new data. GANs can be employed to create realistic three-dimensional models of mineral deposits (Introduction to machine learning, 2019).

When applied to solving practical issues in petroleum geology, the aforementioned ML and neural network methods should consider the specific geological environment in each case. For example, the subsalt carbonate of the Precaspian depression (including its southeastern board) is the primary targets for hydrocarbon production. These sediments are characterized by high heterogeneity and complex structures including widespread fracturing making them one of the most challenging objectives for traditional analysis methods.

Nevertheless, modern ML algorithms, such as SVM, RF, ANN, CNN, RNN, and GAN enable effective solutions to classification, regression, and modeling targets in complex fractured environments.

For instance, in studies of carbonate reservoirs, ANNs were used to predict the distribution of fracturing based on seismic inversion data. The model improves the correlation between geophysical parameters and actual fracturing data obtained from core samples. Of particular importance is the ability to analyze the orientation, density, and size of fractures, which is crucial to comprehend the filtration properties of reservoirs.

Neural Network Learning

The fundamental element of neural networks is the "mathematical neuron" or perceptron. This is a simplified mathematical model of a biological neuron, which forms the basis of ANN (Neural networks for beginners, 2023).

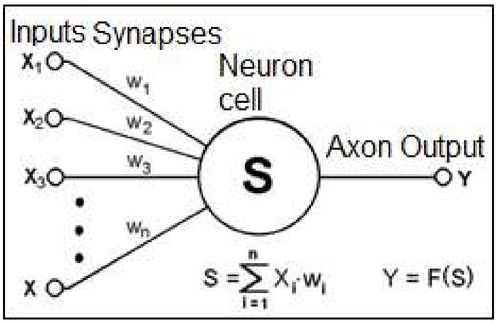

Fig. 1. Artificial neuron (Neural networks for beginners, 2023)

A neuron performs a nonlinear transformation of input signals. First, the values of the input signals (x_i) multiplied by their corresponding weights (w_i). Then, the resulting products were summed. The result of this summation undergoes a nonlinear transformation using an activation function. The output of this transformation is the neuron's output signal (Neural networks. psy.wikireading, 2023).

In Fig. 2 the algorithm of lithofacies and reservoir properties evaluation were shown using Kolmogorov neural networks (Simplified explanation of the new Kolmogorov-Arnold network from MIT, 2024) based on seismic data with the application of deep networks (convolution and shifting are implemented within tensors). The activation function plays an independent role and was defined separately. Typically, this is a sigmoid function, where the perceptron's response can be determined as yes/no based on cutoff value of the activation function.

At the same time, it is important to consider that modern industry software already incorporates ML algorithms for calculating inversion based on seismic data, which produce an inversion cube as output. The primary focus should be on calibrating seismic and well data to enhance the resolution of seismic impedance.

KAN were used to approximate nonlinear relationships between input data (well logs and seismic attributes) and output parameters (lithology and reservoir properties). Cross-validation allows assessing how well the trained network will perform on new, previously unseen data and helps to prevent overfitting.

Fig. 3 shows a cross-plot comparing the results of calculating parameters such as PIMP (P-wave acoustic impedance), VP/VS (ratio of compressional to shear wave velocities), and porosity obtained using two different methods: the synchronous inversion algorithm (red curve) and ML algorithms (blue curve).

The comparison reveals that the results were obtained by the two methods align, but differences are observed, which attributed to the methodologies of the two approaches.

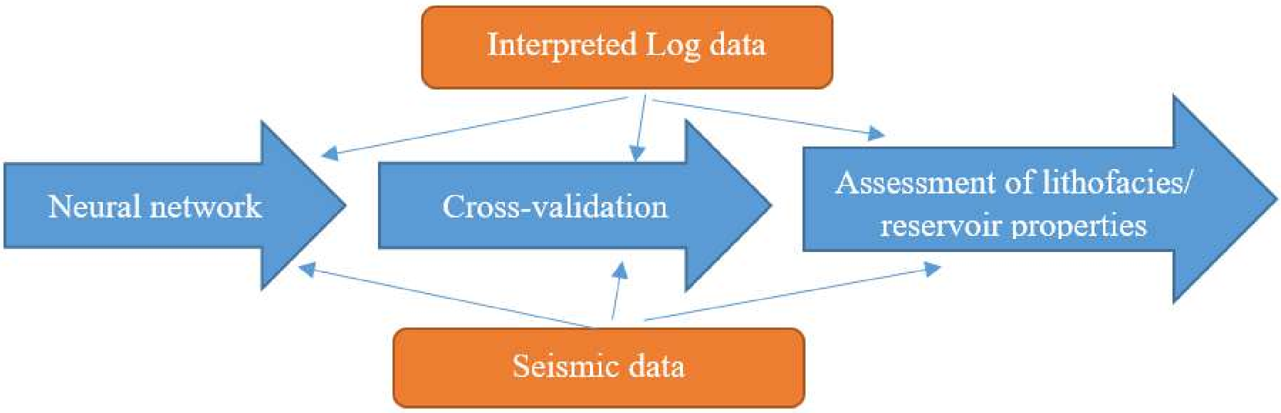

Fig. 2. Evaluation of lithofacies and reservoir properties using KAN (Simplified explanation of the new Kolmogorov-Arnold network from MIT, 2024)

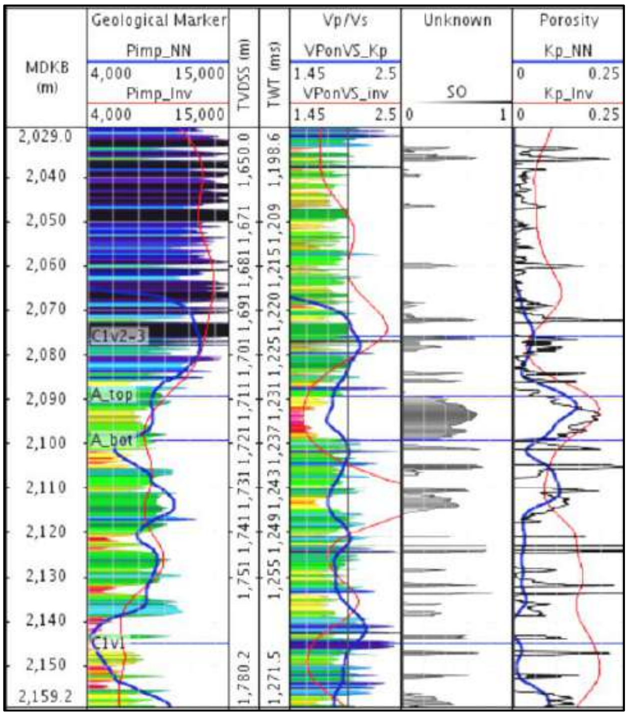

Fig. 3. Comparison of parameters calculated by the synchronous inversion algorithm (red curve) and ML algorithms (blue curve) with the original well logs (based on data from PGSK, 2025)

Synchronous inversion is based on physical models and requires precise knowledge of initial parameters, such as the acoustic properties of rocks and boundary conditions. It can be more sensitive to noise in the data and requires careful calibration.

ML uses data to train models that can identify complex nonlinear relationships not always explicitly described by physical equations. To mitigate these differences, cross-plotting across the entire dataset is necessary.

The article (Приезжев и др., 2023) describes the successful application of neural network prediction (using KAN) for creating geological models of hydrocarbon fields. In particular, upon completion of drilling Well 7, the predicted distribution of the reservoir was confirmed, which allowed refining the 3D model and optimizing drilling, thereby improving the economic efficiency of the project (see Fig. 4) (Приезжев и др., 2023).

A consolidated reservoir (characterized by a high degree of compaction and cementation, leading to the predominance of secondary porosity) identified using neural network prediction (highlighted contour on the upper cross-section), was confirmed by Well 7 (Приезжев и др., 2023).

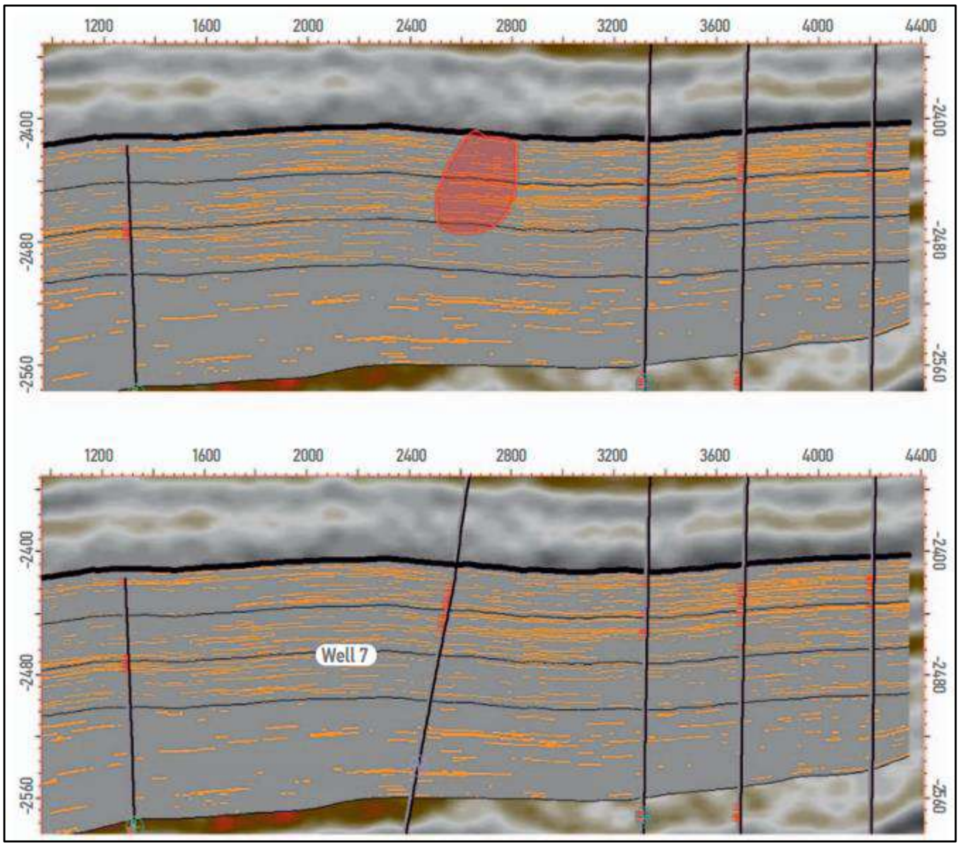

Fig. 4. Cross-section from the three-dimensional reservoir model- extracted using the neural network prediction methodology before drilling Well 7 (top), and the updated model after drilling Well 7 (bottom) (Приезжев и др., 2023)

The possibility of comparing the net pay of the model (H_eff model) and the net pay determined from well logging data (H_eff logging) exists, but it requires a comprehensive approach.

To compare H_eff model and H_eff logging, it is necessary to calibrate the model based on well logging data. This involves comparing calculated values with real data and adjusting model parameters such as porosity, permeability, and saturation. An important step is the construction of petrophysical relationships that associated parameters measured in wells with logging data. For example, the relationship between acoustic impedance and porosity allows refining the model and improving its alignment with real data (Добрынин, 2009).

In the study (Priezzhev et al., 2019), a method for analyzing seismic facies was proposed using 3D Kohonen neural networks. This approach enables the establishment of complex, nonlinear relationships between various parameters of seismic traces. Optimization of the classification process was achieved through the application of ML and self-organizing maps (SOM). RGB blending of 1D, 2D, and 3D projections were used to visualize the results, which helps identify relationships between classes and obtain a more accurate geological interpretation based on well data and analogs (Priezzhev et al., 2019).

RGB blending is a data visualization method in which three different parameters or datasets were displayed as color channels (red, green, and blue) to create the integrated image (Браун, 2011).

According to the article (Колбикова, 2021), the tuning of the cluster model was carried out in several stages which allows solving objectives related to lithotype classification and saturation prediction. Initially, a model was created to group limestones based on their ability to hand over oil. Then, a more detailed model to distinguish various rock types was developed using well logging and core data. Accordingly, the statistical algorithm Multi-Resolution Graph-based Clustering (MRGC) was applied, which associates geophysical data to rock type. This method is based on comparing new data with known examples and takes into account various rock characteristics. As a result, five main rock types including two reservoir types were identified (Колбикова, 2021).

In the paper (Merembayev et al., 2021), the assessment of lithofacies was studied using ML algorithms of geological data from Kazakhstan and Norway. ML methods such as k-nearest neighbors (kNN), decision trees, RF, XGBoost, and Light GBM both with and without wavelet transformation of the data were considered. Input data of the ML models include gamma ray, acoustic and neutron logs, etc. (Merembayev et al., 2021).

For the evaluation of lithofacies and reservoir properties in the southeastern board of the Precaspian Basin, it is advisable to use a comprehensive approach combining various AI methods considering the complexity of the region's geological structure.

For lithofacies classification, where high accuracy and resistance to overfitting are required, methods such as RF are most suitable. SVM and neural network learning are recommended to assess reservoir properties, where the ability to approximate nonlinear relationships is important.

However, any ML method is resistant to overfitting, and this is achieved within the basic "workflow." In this case, the key factor is the balance between accuracy and computation time. The calculations can be performed on different models followed by selecting the one that best solves the problem considering limitations on accuracy and time. To enhance the robustness of the calculations, a "voting" model can be applied~~,~~ incorporating several different models.

For example, as it is noted above, the Upper Paleozoic fractured carbonate reservoirs in the southeastern Precaspian Basin is characterized by high heterogeneity due to the presence of secondary porosity (fractures, vugs, and leaching channels). Analyzing such reservoirs requires a method capable of accounting for spatial variability and identifying lithological and facies features.

The most suitable AI method for solving these targets is CNN, which can effectively process seismic data and well logs, enabling the identification of fracture systems and vuggy intervals that play a key role in the formation of the filtration and storage properties of productive reservoirs.

Based on the analysis of inversion results and neural network learning with actual well data, it was established that neural network learning demonstrates convergence in absolute values and possesses high resolution. This is due to the consideration of nonlinear transformations that neural networks can effectively be modeled. The use of a nonlinear operator based on neural networks and evolutionary algorithms provides the following advantages:

- The model has the ability to establish nonlinear relationships between heterogeneous types of data significantly improving the accuracy of interpretation and prediction of characteristics of set of geological objects, such as fractured carbonate reservoirs, where traditional methods may not account for complex interrelationships between lithological, textural, and structural features of rocks.
- The model automatically adapts to various types of input data and conditions~~,~~ increasing its efficiency.

Automatic tuning of model parameters contributes to achieving optimal results, i.e., obtaining hyperparameters. The search for these defined by an iterative solution search code based on gradient descent – this is part of the workflow in model tuning.

Gradient descent, or the gradient descent method is a numerical method for finding the local minimum or maximum of a function by moving along the gradient, and it is one of the primary numerical methods in modern optimization (Scikit-learn official documentation)

4. Discussion

For the subsalt reservoirs of the southeastern Precaspian Depression, as it is noted above, the most suitable methods for assessing lithofacies and reservoir properties are SVM and RF.

Both methods exhibit high sensitivity to noise, which is crucial for processing geophysical data that usually contain interference. They also have medium computational complexity (Table 1) making them accessible for processing.

An important factor is that both methods are well-adapted for processing, interpreting, and modeling geological sections including seismic sections, which play a key role in studying the Precaspian Depression.

In addition, RF also exhibits resistance to overfitting, which is a significant advantage while working with limited data. Overall, the choice between SVM and RF depends on the specific task and available resources. If high classification accuracy is required, RF may be preferable. Another advantage of RF is that it does not require data standardization.

If speed is important, SVM may be more suitable.

The application of neural network learning in the southeastern board of the Precaspian Depression can demonstrate its effectiveness due to its ability to account for nonlinear relationships between seismic fields and parameters measured in wells.

The main factors contributing to the advantages of neural networks include:

- Accounting for nonlinear distortions: Neural networks can model nonlinear dependencies. This is particularly important in geological settings where salt structures, fault zones, and other distorting factors are present.
- Absence of theoretical limitations: Neural networks do not require strict theoretical frameworks to handle nonlinear distortions making them more versatile.

However, neural networks are just one group among many ML methods and do not always yield the best results. In such cases, it is necessary to adopt a comprehensive approach, i.e., seismic data and well measurements to build a high-frequency model, which significantly enhances resolution and prediction accuracy.

The tasks addressed using AI in lithofacies analysis can be divided into several main categories, as presented in Table 2.

Table 1

Comparison of characteristics of AI methods (Shamaev, 2022)

AI method name	Computational complexity	Amount of data required for training	Sensitivity to noise	Suitability for processing 2D geophysical data
SVM	Medium	Small	High	Medium
Relevance Vector Machines	High	Small	Low	Medium
Decision Tree	Medium	Medium	High	Medium
RF	Medium	Medium	High	Medium
Genetic Algorithms	High	Not Required	Low	Poor
Gaussian Process Regression	Medium	Small	Medium	Poor
kNN	Low	Medium	Low	Poor
Logistic Regression	Medium	Medium	Low	Poor
Ridge Regression	Medium	Medium	Medium	Medium
Jackknife Regression	Medium	Medium	Medium	Medium
ANN (CNN, RNN, GAN)	High	Medium	Low	Good

Table 2

Categories of tasks solving using AI

Task/Function	Objective	Methods
Facies / lithofacies classification	Assigning rocks to specific facies types based on available data (lithological, geophysical, petrophysical).	• SVM – effective for classifying complex data, handles nonlinear class boundaries well. • RF – ensemble of decision trees, resistant to overfitting and noise, provides good classification accuracy. • Naive Bayes Classifier – simple and fast method, works well on data with independent features.
Regression for reservoir property estimation	Estimating quantitative reservoir properties (porosity, permeability, oil saturation) based on available data.	• Linear Regression – method used to establish a linear relationship between variables. • Polynomial Regression - An extension of linear regression, allowing modeling of nonlinear relationships.
Clustering	Dividing a dataset into groups (clusters) based on similarity of their characteristics.	• k-means – simple and popular method that divides data into k clusters by minimizing intra-cluster distance. • Hierarchical Clustering – Builds a hierarchy of clusters, allowing analysis of data structure at different levels.
Image segmentation	Identifying regions in images corresponding to different minerals, pores, fractures, and other elements.	• CNN: Architectures like U-Net, SggNet, and others – Show high effectiveness in image segmentation tasks.

Thus, the application of AI methods for lithofacies analysis and reservoir property assessment in the southeastern Precaspian Basin opens new opportunities for more accurate and efficient interpretation of geological, geophysical, and field data. However, the choice of a specific AI method requires careful analysis and consideration of several factors (accuracy, data completeness, and desired results).

6. Conclusions

The article analyzes the application of AI methods in solving objectives related to lithofacies mapping and reservoir property assessment of play zones, which is a highly promising direction capable of significantly improving the efficiency and accuracy of geological exploration. This facilitated by the diversity of AI methods, such as ML, neural networks, and genetic algorithms, which provide a wide range of tools for solving various objectives related to the analysis of geological, geophysical, and field data.

For facies and lithofacies classification, methods such as SVM, RF, and the Naive Bayes classifier are recommended. Regression methods, including linear and polynomial regression can be applied for reservoir property assessment.

A key aspect for ML methods is understanding the parameters to be worked with – whether they are continuous or discrete – and whether the applied algorithms have versions for both regression and classification.

Clustering, particularly k-means and hierarchical clustering also finds applications in geology. For image segmentation, CNN are used, although the option of using other tools is not excluded.

Practical examples show that CNNs are effective for identifying fractures in rock masses due to their ability to process 2D geophysical data (Шамаев, 2022). For lithofacies classification goals based on well logging data, KAN can be useful due to their improved approximation and interpretability (Simplified explanation of the new KAN from MIT, 2024).

At the same time, it is necessary to check the sensitivity of models to their inputs and evaluate the importance of each variable in the optimized model. This verification was conducted using a separate library, which enhances the interpretability of most models.

For seismic facies analysis, effective methods such as 3D Kohonen neural networks were proposed, which allow establishing complex, nonlinear relationships between various parameters of seismic traces. Optimization of the classification process was achieved through the application of ML and SOM. Visualization of results using RGB blending of 1D, 2D, and 3D projections helps to identify relationships between classes and obtain a more accurate geological interpretation based on well data and analogs.

In summary, the most effective methods were recognized as SVM, RF, and neural networks for the southeastern board of Precaspian Basin.

- RF exhibits high resistance to overfitting and can handle high-dimensional data~~,~~ making it useful for classification and prediction.

- Neural networks demonstrate high accuracy and resolution due to their ability to account for nonlinear distortions in seismic signals and dependencies between geophysical parameters.
- SVM have proven effective in lithofacies classification goals, especially in conditions of limited data volume and the presence of noise.

At the same time, it is important to note the integration of several AI methods, which enhances the accuracy of predicting reservoir properties and lithofacies.

Despite the obvious successes achieved, the application of AI in geology and geophysics remains a relevant and promising direction with many issues requiring further study, understanding, and handling. The problems related to the interpretation of results, big data, training data, and the adaptation of existing methods to specific geological conditions are among these.

Nevertheless, it can be confidently stated that AI will continue to play an increasingly important role in geology contributing to the development of new approaches to subsurface exploration and providing more accurate and efficient solutions to objectives related to the search and exploration of mineral resources.

REFERENCES

Brown A.R. Interpretation of Three-Dimensional Seismic Data. Translated from English. Society of Exploration Geophysicists (SEG). Moscow, 2011 (in Russian).
Dobrynin V.M., Vendelstein B.Yu., Kozhevnikov D.A. Interpretation and modeling of geophysical data. Nedra. Moscow, 2009, 456 p. (in Russian).
Easyoffer.ru. What is model overfitting. URL: https://easyoffer.ru/question/5440 (accessed: 10.10.2023).
Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: Data mining, inference, and prediction. 2nd ed. Springer. New York, 2009, 745 p.
Introduction to machine learning. URL: https://habr.com/ru/post/448892/ (accessed: 3.09.2019) (in Russian).
Keras: The Python Deep Learning API. Official Documentation. URL: https://keras.io/.
Kolbikova E.S. lithofacies analysis and prediction of properties based on geophysical and seismic data using machine learning methods.: Roxar Paradigm – Software and Solutions LLC. Moscow, 2021, 120 p. (in Russian).
KPMG Caucasus and Central Asia. Technological Research 2024. KPMG. February 2025.
Merembayev T., Kurmangaliyev B., Bekbauov B., Amanbek Ye. A comparison of machine learning algorithms in predicting lithofacies: Case studies from Norway and Kazakhstan. Energies. Vol. 14, No. 7, 2021, p. 1896, DOI: https://doi.org/10.3390/en14071896.
Neural networks for beginners. Part 1. Habr. URL: https://habr.com/ru/articles/312450 (accessed: 10.10.2023) (in Russian).
Neural networks. Psy.Wikireading. URL: https://psy.wikireading.ru/12215 (accessed: 10.10.2023) (in Russian).
Ng A. Artificial Intelligence (AI) for Everyone. Deep Learning.AI. 2023, URL: https://www.deeplearning.ai (accessed: 10.10.2023).
Priezzhev I.I., Taikulakov E E., Kayumov I. L., Leonov A.V., Gorbach D.A., Miroshkin V.G., Ovechkina V.Yu. Direct neural network prediction of reservoir properties based on seismic data: A case study of clinoform deposits in Western Siberia. PRONeft. Professionalno o nefti, Vol. 8, No. 2, 2023, pp. 28-39, DOI: https://doi.org/10.51890/2587-7399-2023-8-2-28-39 (in Russian).
Priezzhev I. I., Veeken P. C. H., Egorov S.V., Nikiforov A.N., Strecker U. Seismic waveform classification based on Kohonen 3D neural networks with RGB visualization. First Break, Vol. 37, No. 2, 2019, pp. 37-43.
Scikit-learn: Machine Learning in Python. Official Documentation. URL: https://scikit-learn.org/stable/user_guide.html.
Shamaev S.D. Application of artificial intelligence methods in processing and interpretation of geophysical data. Izvestiya UGGU, No. 1 (65), 2022, pp. 86-101 (in Russian).

ЛИТЕРАТУРА

Easyoffer.ru. Что такое переобучение модели. URL: https://easyoffer.ru/question/5440 (дата обращения: 10.10.2023).
Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: Data mining, inference, and prediction. 2nd ed. Springer. New York, 2009, 745 p.
Merembayev T., Kurmangaliyev B., Bekbauov B., Amanbek, Ye. A comparison of machine learning algorithms in predicting lithofacies: Case studies from Norway and Kazakhstan. Energies. Vol. 14, No. 7, 2021, p. 1896, DOI: https://doi.org/10.3390/en14071896.
Ng A. Artificial Intelligence (AI) for Everyone. Deep Learning.AI. 2023, URL: https://www.deeplearning.ai (дата обращения: 10.10.2023).
Priezzhev I.I., Veeken P.C.H., Egorov S.V., Nikiforov A.N., Strecker U. Seismic waveform classification based on Kohonen 3D neural networks with RGB visualization. First Break, Vol. 37, No. 2, 2019, pp. 37-43.
Scikit-learn: Machine Learning in Python. Официальная документация. URL: https://scikitlearn.org/stable/user_guide.html.
Браун А.Р. Интерпретация трехмерных сейсмических данных. Пер. с англ. Общество инженеров-геофизиков (SEG), Москва, 2011.
Введение в машинное обучение. URL: https://habr.com/ru/post/448892/ (дата обращения: 3.09.2019).
Добрынин В.М., Вендельштейн Б.Ю., Кожевников Д.А. Интерпретация и моделирование геофизических данных. Недра. Moscow, 2009, 456 с.
Keras: The Python Deep Learning API. Официальная документация. URL: https://keras.io/.
Колбикова Е.С. Литофациальный анализ и возможности прогнозирования свойств по данным геофизических исследований и сейсморазведки методами машинного обучения. ООО «Роксар Парадайм – ПО и Решения». Москва, 2021, 120 с.
KPMG Кавказ и Центральная Азия. Технологическое исследование 2024. KPMG. Февраль 2025.
Нейронные сети для начинающих. Часть 1. Habr. URL: https://habr.com/ru/articles/312450 (дата обращения: 10.10.2023).
Нейронные сети. Psy.Wikireading. URL: https://psy.wikireading.ru/12215 (дата обращения: 10.10.2023).
Приезжев И.И., Тайкулаков Е.Е., Каюмов И.Л., Леонов А.В., Горбач Д.А., Мирошкин В.Г., Овечкина В.Ю. Прямой нейросетевой прогноз коллекторских свойств пласта по данным сейсморазведки на примере клиноформных отложений Западной Сибири. PROНефть. Профессионально о нефти, Т. 8, No. 2, 2023, с. 28-39, DOI: https://doi.org/10.51890/2587-7399-2023-8-2-28-39.
Шамаев С.Д. Применение методов искусственного интеллекта при обработке и интерпретации данных геофизических методов. Известия УГГУ, № 1 (65), 2022, с. 86-101.

Simplified explanation of the new Kolmogorov-Arnold network from MIT. Habr. 2024. URL: https://habr.com/ru/articles/817547 (accessed: 10.10.2023) (in Russian).

Упрощённое объяснение новой сети Колмогорова-Арнольда из MIT. Habr. 2024. URL: https://habr.com/ru/articles/817547 (дата обращения: 10.10.2023).

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ И НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ КАРТИРОВАНИЯ ЛИТОФАЦИЙ И ОЦЕНКИ КОЛЛЕКТОРСКИХ СВОЙСТВ: АНАЛИЗ И ВЫБОР МЕТОДОВ

Абетов А.Е.[1], Сейтжанов А.К.[2], Саменов Е.Р.[1]

[1]Казахский национальный исследовательский технический университет имени К. И. Сатпаева, Казахстан
Ул. Сатпаева 22, Алматы, 050013
[2]Казахстанско-Британский Технический Университет, Казахстан
Ул. Толеби, 59, Алматы, 050000: ansar.seitzhanov.98@mail.ru

Резюме. В статье рассмотрены аспекты применения методов искусственного интеллекта (ИИ) для решения задач картирования литофаций и оценки коллекторских свойств пород. При этом выбор метода ИИ зависит от характера данных, целей исследования (классификация, регрессия, кластеризация, сегментация изображений), требований к конечным результатам интерпретации и моделирования. Проведен анализ различных алгоритмов машинного обучения, таких как метод опорных векторов, случайный лес, нейронные сети и другие. Оценена эффективность каждого их перечисленных методов на основе открытых источников и геологических данных. Проанализированы их преимущества и недостатки, а также идентифицированы факторы, влияющие на выбор метода ИИ. Представлена классификация геологических задач и соответствующих методов ИИ, включая SVM, RF, линейную и полиномиальную регрессию, k-means, иерархическую кластеризацию и CNN. Представлены платформы машинного обучения с открытым исходным кодом, такие как TensorFlow, PyTorch и Keras, а также факторы, влияющие на выбор оптимального метода ИИ для литофациального анализа и оценки коллекторских свойств. Предложены рекомендации по выбору оптимальных методов ИИ для решения конкретных задач. Подчеркнута важность объема и качества данных для выбора метода ИИ и предотвращения переобучения модели. Наибольшую результативность в условиях Юго-Восточного Прикаспия могут продемонстрировать метод опорных векторов, случайный лес и нейронные сети, каждый из которых обладает уникальными преимуществами. В работе отмечены ключевые проблемы, связанные с качеством исходных данных, интерпретируемостью результатов и адаптацией методов к конкретным геологическим условиям, а также переобучение. В обсуждении и заключении представлена информация касательно применения каждого метода машинного обучения под конкретные задачи и условия.

Ключевые слова: *искусственный интеллект, машинное обучение, классификация геологических задач, линейная и полиномиальная регрессия, кластеризация литофаций, коллекторские свойства*

LİTOFASİYALARIN XƏRİTƏLƏNDİRİLMƏSİ VƏ KOLLEKTOR XÜSUSİYYƏTLƏRİNİN QİYMƏTLƏNDİRİLMƏSİ TAPŞIRIQLARINDA MAŞIN ÖYRƏNMƏSİ VƏ NEYRON ŞƏBƏKƏ METODLARININ TƏTBİQİ: METODLARIN TƏHLİLİ VƏ SEÇİMİ

Abetov A.E.[1], Seyidzhanov A.K.[2], Samenov E.R.[1]

[1]K. İ. Satpayev adına Qazax Milli Tədqiqat Texniki Universiteti, Qazaxıstan
Satpayev küç., 22, Almatı, 050013
[2]Qazaxıstan-Britaniya Texniki Universiteti, Qazaxıstan
Tolebi küç., 59, Almatı, 050000: ansar.seitzhanov.98@mail.ru

Xülasə. Məqalədə süni intellekt (Sİ) metodlarının litofasiyaların xəritələşdirilməsi və süxurların kollektor xüsusiyyətlərinin qiymətləndirilməsində tətbiq imkanları araşdırılmışdır. Sİ metodunun seçimi məlumatların xarakterindən, tədqiqatın məqsədlərindən (təsnifat, reqressiya, klasterləşmə, təsvirlərin seqmentasiyası), eləcə də interpretasiya və modelləşdirməyə qoyulan tələblərdən asılı olaraq müəyyən edilir. SVM, RF və CNN və digər maşın öyrənməsi alqoritmləri təhlil olunmuşdur. Bu metodların effektivliyi açıq mənbəli məlumatlar və real geoloji verilənlər əsasında qiymətləndirilmiş, onların üstünlükləri və məhdudiyyətləri müəyyən edilmişdir. Süni intellekt metodlarının seçiminə təsir edən əsas amillər müəyyənləşdirilmiş və geoloji vəzifələrin təsnifatına uyğun metodlar təqdim edilmişdir. Bunlara SVM, RF, xətti və polinomial reqressiya, k-means klasterləşməsi, iyerarxik klasterləşmə və konvolyusion neyron şəbəkələr (CNN) daxildir. TensorFlow, PyTorch və Keras kimi açıq mənbə kodlu maşın öyrənməsi platformaları təqdim olunmuş və litofasiyaların analizi ilə kollektor xüsusiyyətlərinin qiymətləndirilməsi üçün optimal metodun seçiminə təsir göstərən amillər müzakirə edilmişdir. Müəyyən geoloji məsələlərin həlli üçün konkret metodlara dair tövsiyələr verilmişdir. Modelin həddindən artıq öyrədilməsi (overfitting) riskinin qarşısının alınması məqsədilə verilənlərin həcmi və keyfiyyətinin əhəmiyyəti xüsusi vurğulanmışdır. Cənub-Şərqi Xəzər regionunun geoloji şəraitində dəstək vektorları metodu, təsadüfi meşə və neyron şəbəkələrin yüksək nəticə verdiyi müəyyən edilmişdir. Hər bir metodun spesifik üstünlükləri qeyd olunmuşdur. Araşdırmada ilkin məlumatların keyfiyyəti, nəticələrin interpretasiya imkanları və metodların konkret geoloji şəraitə uyğunlaşdırılması ilə yanaşı, modelin həddindən artıq öyrədilməsi ilə bağlı əsas problemlər də nəzərdən keçirilmişdir. Müzakirə və nəticələr bölməsində isə maşın öyrənməsi metodlarının konkret geoloji məsələlər və onların tətbiq oluna biləcəyi şəraitə uyğun istifadəsi ətraflı təhlil olunmuşdur.

Açar sözlər: *süni intellekt, maşın öyrənməsi, geoloji vəzifələrin təsnifatı, xətti və polinomial reqressiya, litofasiyaların klasterləşdirilməsi, kollektor xüsusiyyətləri*