



SUPPLEMENTARY INFORMATION FOR OTDK THESIS

EVALUATION OF SHALE RESERVOIRS BY ESTIMATING TOC, MINERAL
COMPOSITION, AND POROSITY USING WELL LOGS

Prepared by:

Martins Uchenna Obidiegwu

NEPTUN CODE: ROGF8V

Supervised by:

Prof. Dr. Norbert Péter Szabó

30th December 2024

The supplementary information contains Tables, Plots, Histograms, Maps, Lists of Abbreviations, and Own Results Summary as an Appendix, which has been cited in the original OTDK thesis work for reference.

Appendix I Tables of Correlation Matrix and Descriptive Statistics

Table 4.1 Correlation matrix of input and output parameters in Well 1

	AC	GR	DEN	RD	RS	CN	PE	Clay	Feldspar	Quartz	Porosity
Acoustic log (AC)	1										
Gamma log (GR)	-0.86	1									
Density log (DEN)	0.908	-0.912	1								
Deep Resistivity (RD)	0.814	-0.682	0.87	1							
Shallow Resistivity (RS)	0.272	-0.192	0.261	0.176	1						
Compensated Neutron (CN)	-0.093	0.045	0.027	-0.042	0.152	1					
Photoelectric Factor log (PE)	-0.112	0.05	0.026	-0.032	0.147	0.995	1				
Clay	0.114	-0.058	0.026	0.255	-0.654	-0.424	-0.415	1			
Feldspar	0.041	-0.329	0.13	-0.256	0.307	0.387	0.386	-0.601	1		
Quartz	0.074	0.196	0.006	0.381	-0.321	-0.382	-0.378	0.655	-0.977	1	
Porosity	-0.269	0.127	-0.249	-0.439	0.159	-0.075	-0.054	-0.152	0.359	-0.399	1

After calculating Pearson's correlation coefficient, Table 4.2 shows the summary of statistical parameters for the input and output parameters. Low standard deviation values of the input parameters reflect data points clustered around the mean values. AC, GR, RD, RS, CN, clay feldspar data points, quartz mineral content, and porosity are widely spread, as reflected by the relatively high standard deviation. The data points of AC, GR, RD, RS, CN, clay feldspar, quartz, and porosity are spread out more and widely distributed with a relatively high coefficient of variance.

Table 4.2: Descriptive statistics of input and output parameters in Well 1

Parameters	Minimum	Maximum	Mean	Mode	Median	Standard Deviation	Kurtosis	Skewness
Input	AC	62.38	107.12	77.07	-	72.43	14.72	-0.21
	GR	95.88	450.71	183.9	-	145.05	103.74	1.79
	DEN	2.29	2.62	2.49	-	2.51	0.06	-1.01
	RD	34.4	135.57	75.7	-	74.8	23.61	-0.14
	RS	41.91	144.9	87.64	-	83.84	23.48	-0.43
	CN	12	33.83	19.62	-	18.6	6.01	-0.64
	PE	2.12	3.21	2.67	-	2.69	0.19	-74
Output	Clay	2.68	15.38	8.62	7.96	8.39	2.53	-0.11
	Feldspar	21.39	66.06	51.31	-	53.65	10.24	-1.39
	Quartz	25.52	59.68	32.26	-	28.75	8.27	-0.25
	Porosity	2.9	16.2	8.69	6.1	8.2	2.71	-0.18

The predicted values for the new output mineral composition and porosity of the shale reservoir were obtained, and an equation was generated. These new values for the predicted mineral composition and porosity are now used to make cross-plots to visualize the relationship between the predicted mineral compositions and porosity and the actual variables of porosity and the mineral compositions. The cross-plots in the work were also used to check for outliers, anomalies, and multicollinearity.

Appendix II Statistics Description for Input and Output Features

Table 5.1 Statistical description of all the input and output features (clay) for building the clay model for the ML approach.

	Count	Mean	STD	MIN	25%	50%	75%	MAX
Depth	102	1944.89	322.49	1476.32	1504.24	2082.08	2085.28	2396.60
AC	102	233.62	28.90	181.32	213.11	221.46	249.19	299.14
GR	102	136.03	47.87	67.45	102.48	116.46	157.28	239.17
RILD	102	63.49	26.11	24.28	40.45	58.87	79.18	129.58
DEN	102	809.62	1199.69	2.30	2.51	2.57	2561.12	2603.84
NPHI	102	13.39	10.16	0.15	0.18	15.59	19.62	33.83
Clay	102	9.55	3.91	2.68	7.34	9.24	11.90	17.90

Table 5.2 Statistical description of all the input and output features (feldspar) for building the feldspar model for the ML approach.

	Count	Mean	STD	MIN	25%	50%	75%	MAX
Depth	102	1944.89	322.49	1476.32	1504.24	2082.08	2085.28	2396.60
AC	102	233.62	28.90	181.32	213.11	221.46	249.19	299.14
GR	102	136.03	47.87	67.45	102.48	116.45	157.28	239.17
RILD	102	63.49	26.11	24.29	40.44	58.87	79.18	129.58
DEN	102	809.62	1199.69	2.29	2.50	2.57	2561.12	2603.84
NPHI	102	13.39	10.16	0.15	0.18	15.59	19.62	33.83
Feldspar	102	51.49	12.61	21.39	42.49	54.02	62.20	67.00

Table 5.3 Statistical description of all the input and output features (quartz) for building the quartz model for the ML approach.

	Count	Mean	STD	MIN	25%	50%	75%	MAX
Depth	102	1944.89	322.48	1476.32	1504.24	2082.08	2085.28	2396.60
AC	102	233.62	28.90	181.32	213.11	221.45	249.19	299.14
GR	102	136.03	47.87	67.45	102.48	116.45	157.28	239.17
RILD	102	63.49	26.10	24.28	40.45	58.87	79.18	129.57
DEN	102	809.61	1199.69	2.29	2.51	2.56	2561.12	2603.84
NPHI	102	13.39	10.16	0.15	0.18	15.59	19.62	33.83
Quartz	102	32.21	6.98	23.82	27.21	29.16	35.66	47.77

Table 5.4 Statistical description of all the input features and output features (porosity)

	Count	Mean	STD	MIN	25%	50%	75%	MAX
Depth	102	1944.89	322.48	1476.32	1504.23	2082.08	2085.28	2396.60
AC	102	233.61	28.90	181.32	213.11	221.46	249.19	299.14
GR	102	136.03	47.87	67.45	102.47	116.46	157.28	239.17
RILD	102	63.49	26.10	24.28	40.45	58.87	79.18	129.58
DEN	102	809.61	1199.69	2.29	2.51	2.57	2561.12	2603.84
NPHI	102	13.39	10.16	0.15	0.18	15.59	19.62	33.83
Porosity	102	6.40	3.83	0.10	3.60	6.30	9.73	16.20

Appendix III Histogram of the Input and Output Features

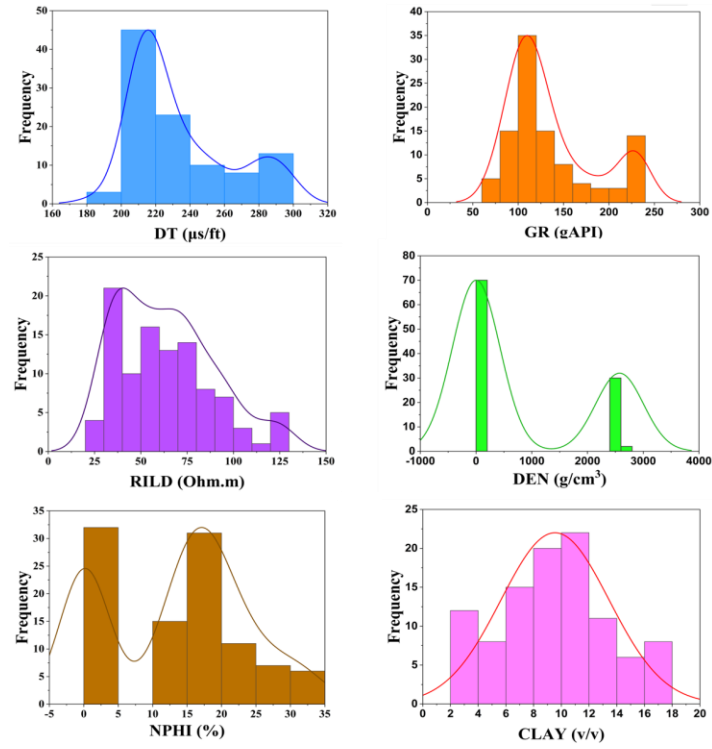


Fig. 5.1: Histograms of the input features and clay volume

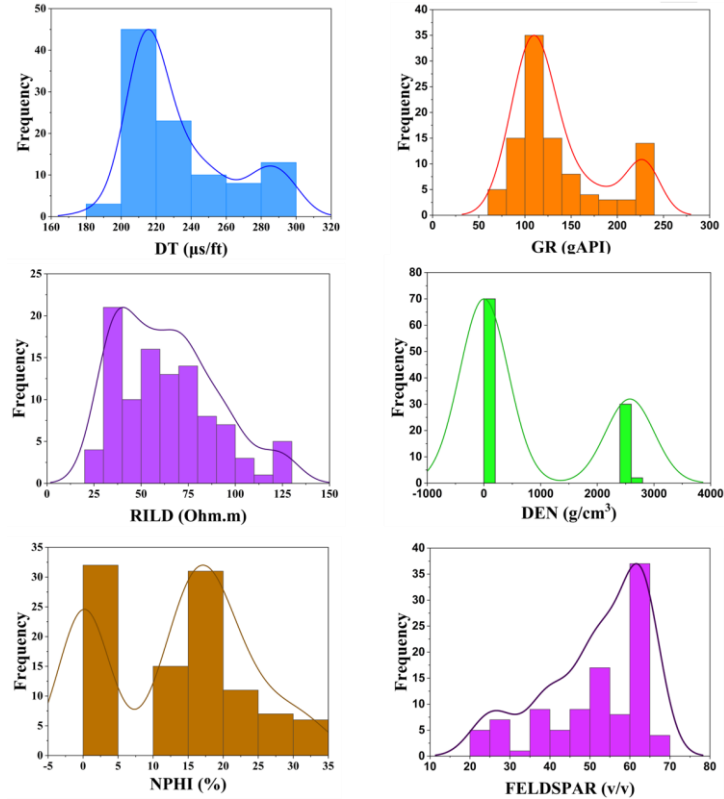


Fig. 5.2: Histograms of the input features and feldspar volume

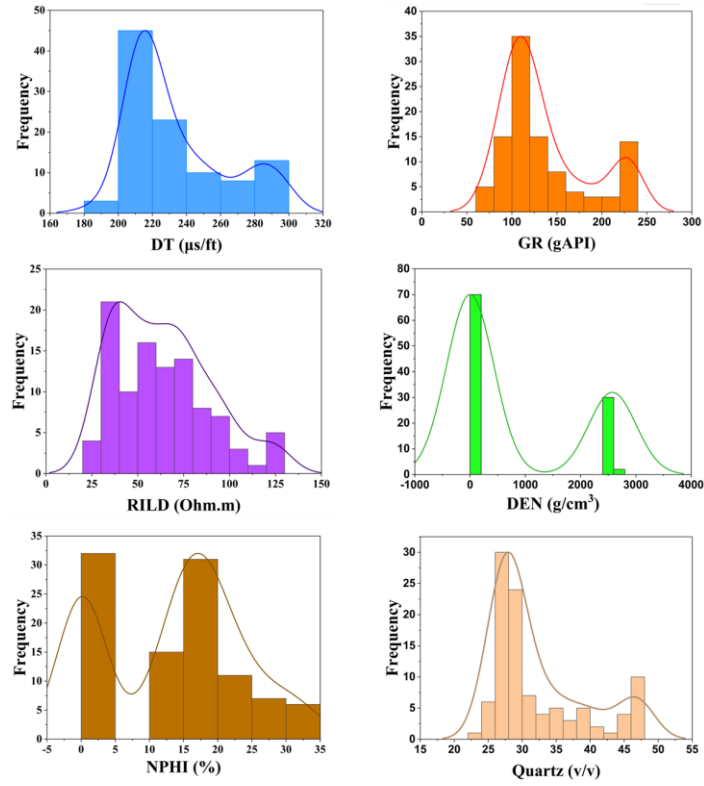


Fig. 5.3: Histograms of the input features and quartz volume

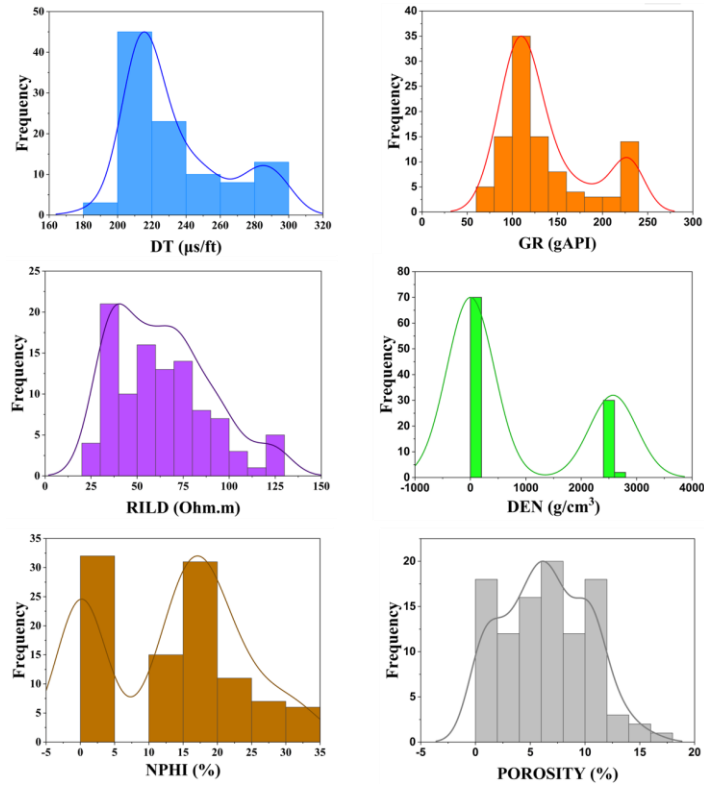


Fig. 5.4: Histograms of the input features and porosity

Appendix IV Box Plots for Input and Output Features

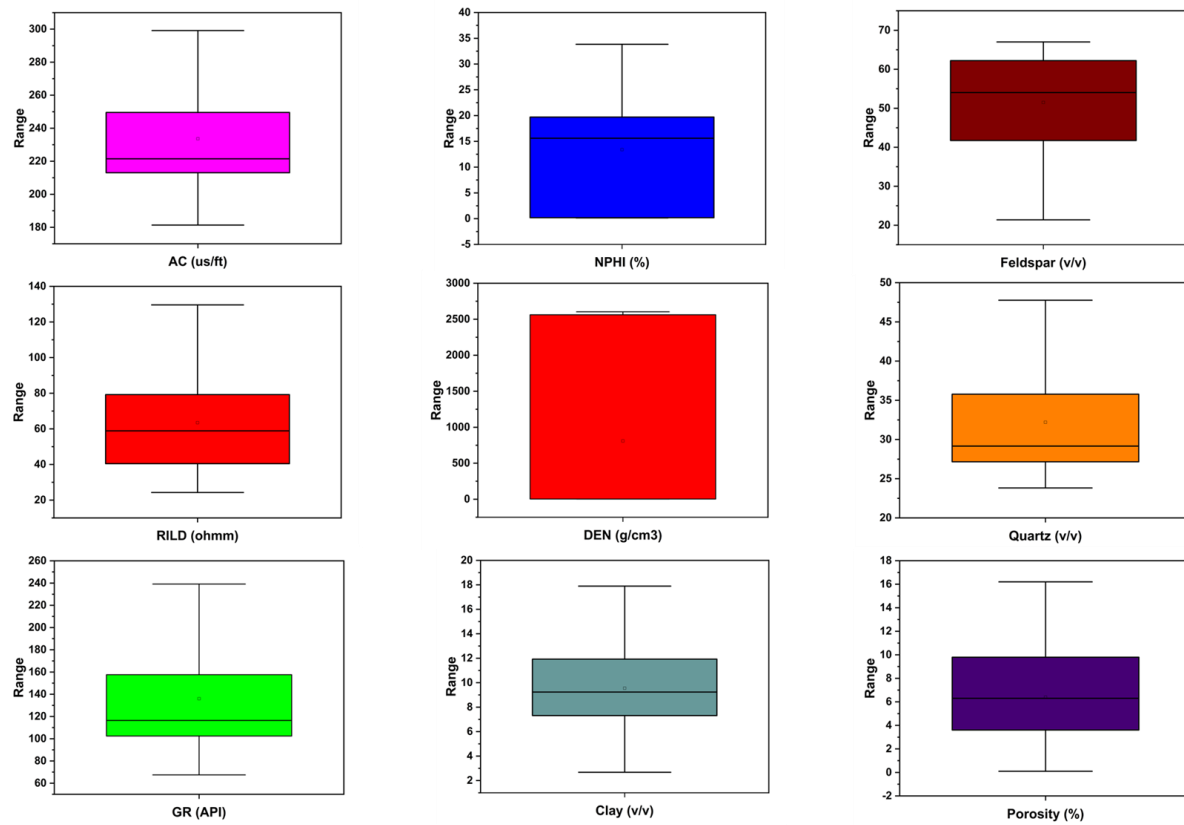


Fig. 5.5: Box plots of the input features and the derived output features (clay content, feldspar volume, quartz volume, and porosity)

Appendix V Heat Map Visualization for Input and Output Features

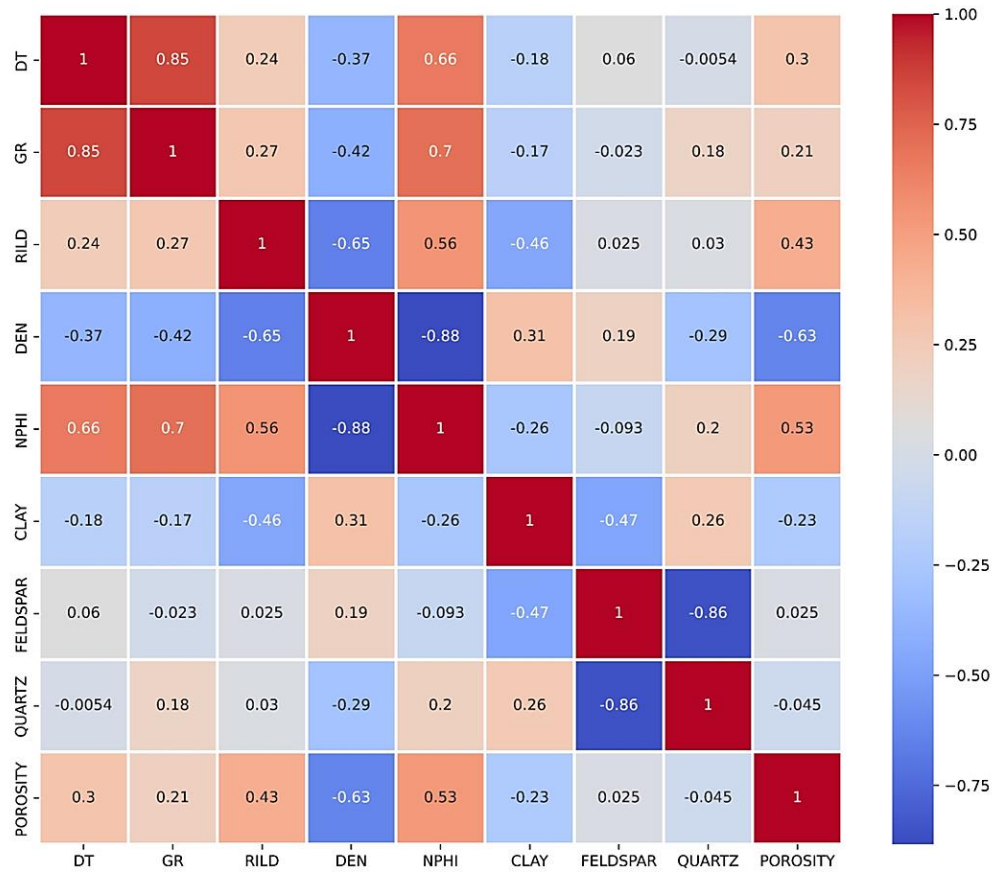


Fig. 5.6: Heat map visualization showing all the input and output petrophysical features.

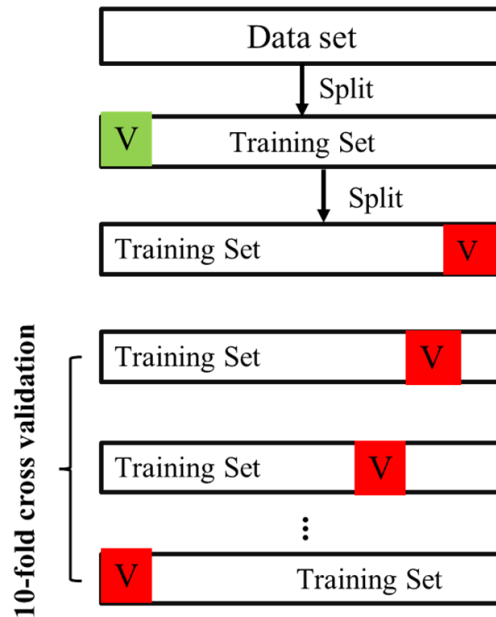


Fig. 5.7: Ten-fold cross-validation flow chart (T is the test data set, V is the validation data set).

Appendix VI Feature and Permutation Importance of Input Parameters and their Outputs from Gradient Boost

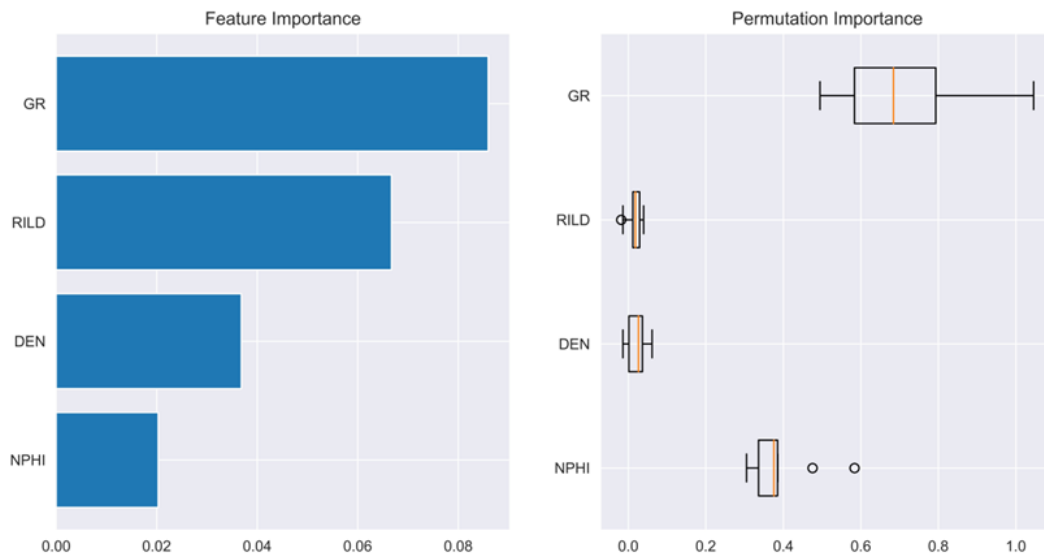


Fig. 5.12: Feature and permutation importance of parameters for clay volume.

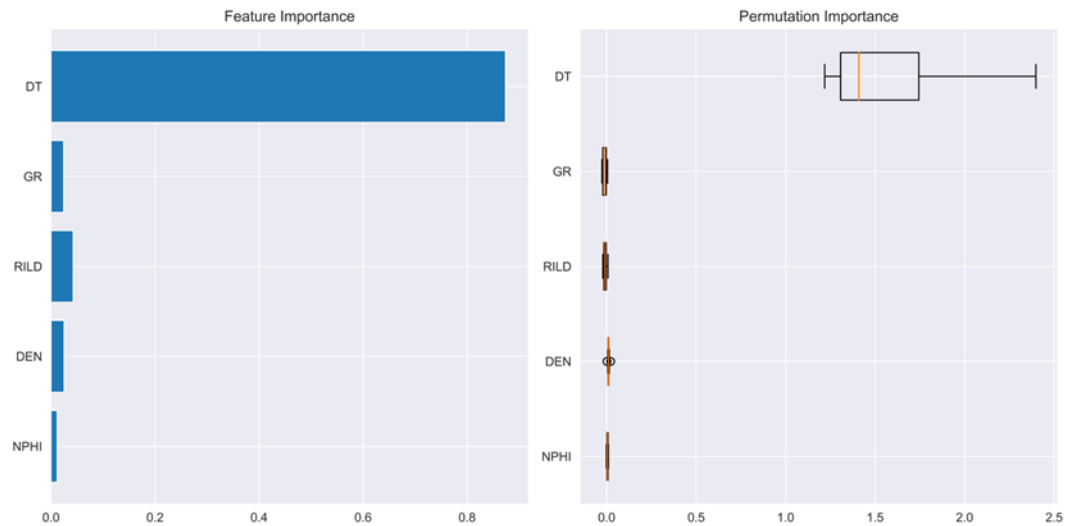


Fig. 5.13: Feature and permutation importance of feldspar volume characteristics.

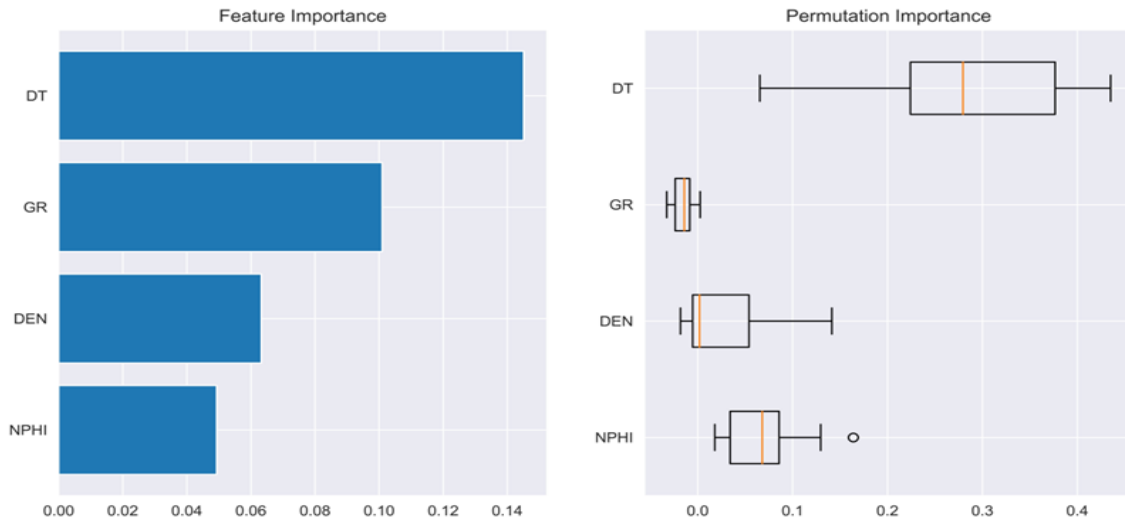


Fig. 5.14: Feature and permutation importance of characteristic parameters for quartz volume.

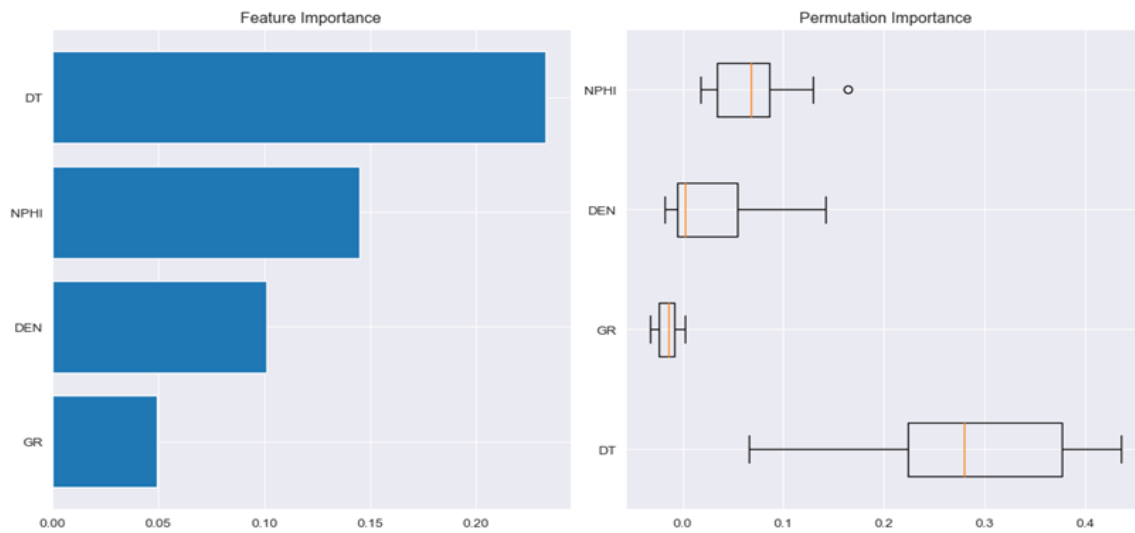


Fig. 5.15: Feature and permutation importance of characteristic parameters for porosity

Appendix VII Feature and Permutation Importance of Input Parameters and their Outputs from XGBoost

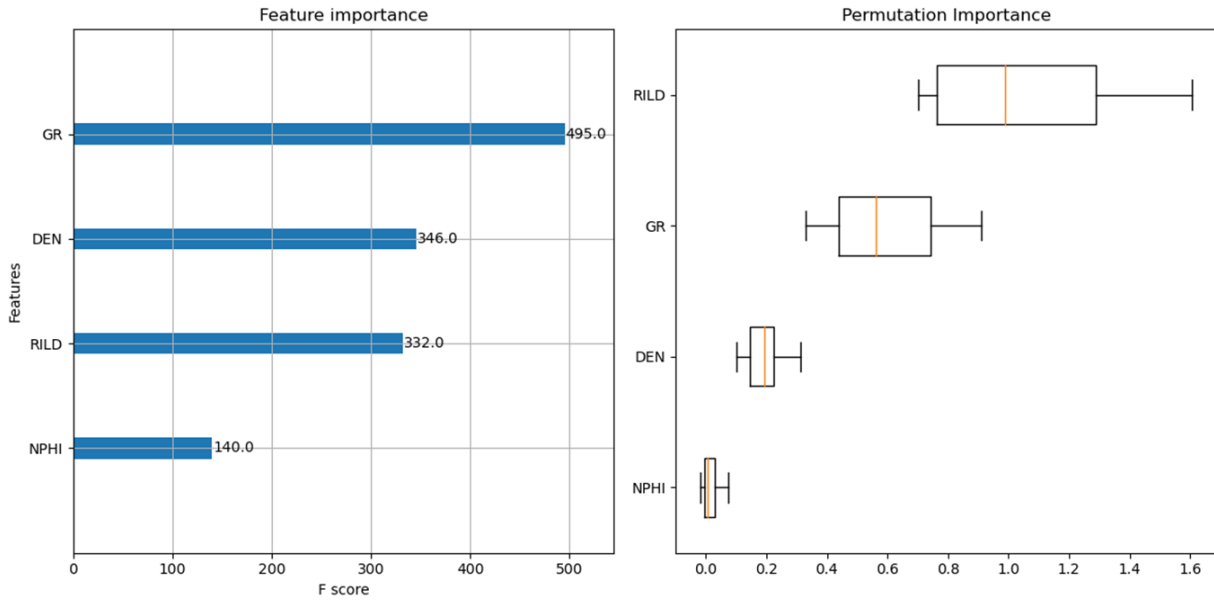


Fig. 5.21: Feature and permutation importance of characteristic parameters for clay volume.

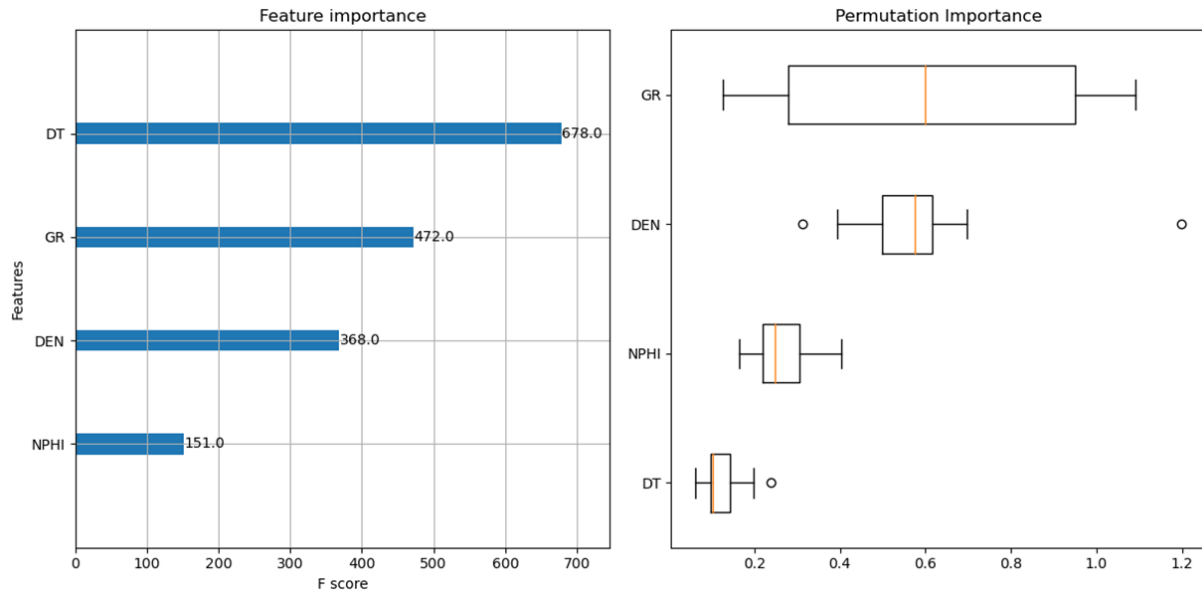


Fig.5.22: Feature and permutation importance of feldspar volume characteristics.

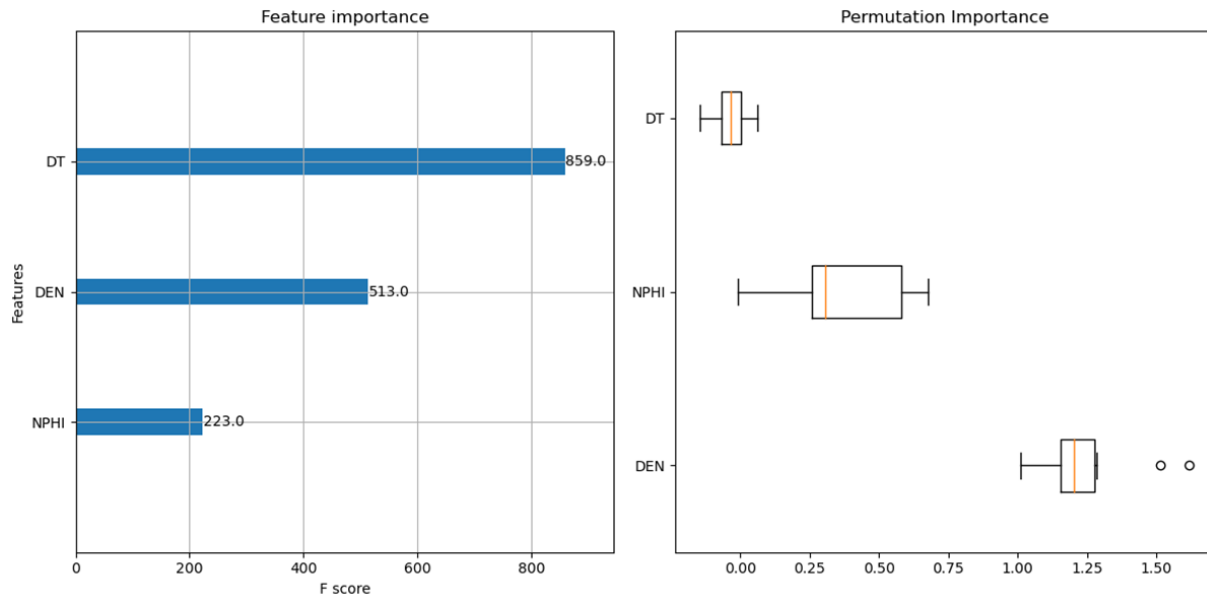


Fig. 5.23: Feature and permutation importance of characteristic parameters for quartz volume.

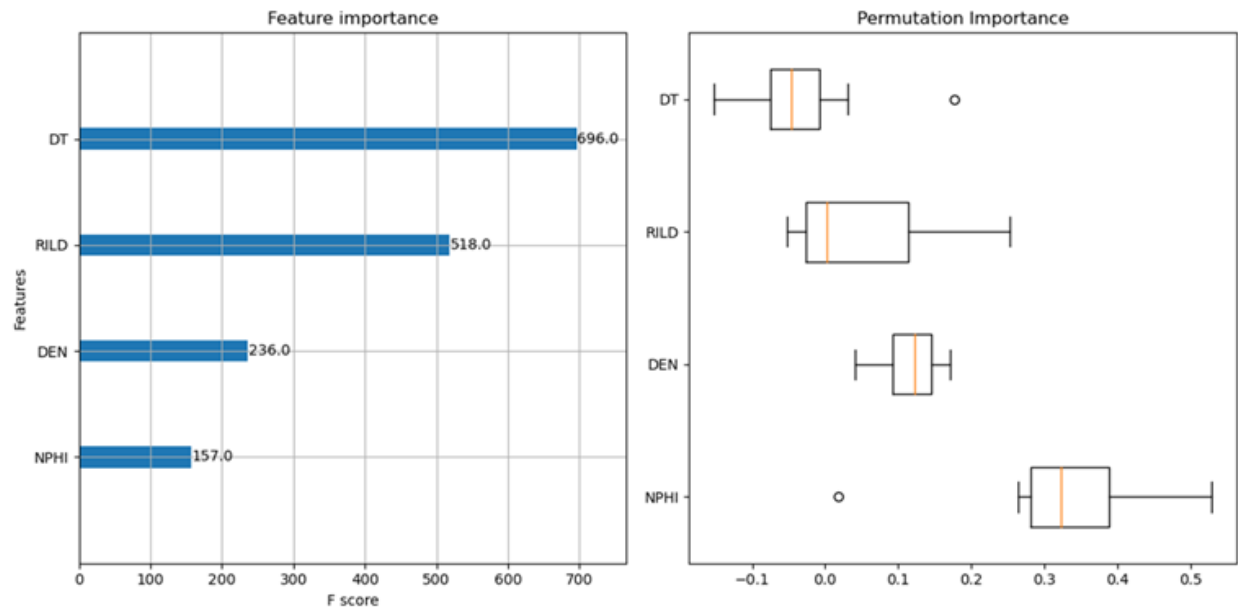


Fig.5.24: Feature and permutation importance of characteristic parameters for porosity.

Appendix VIII Estimation Error Comparison of Training and Testing Results for Clay, Feldspar, Quartz, and Porosity

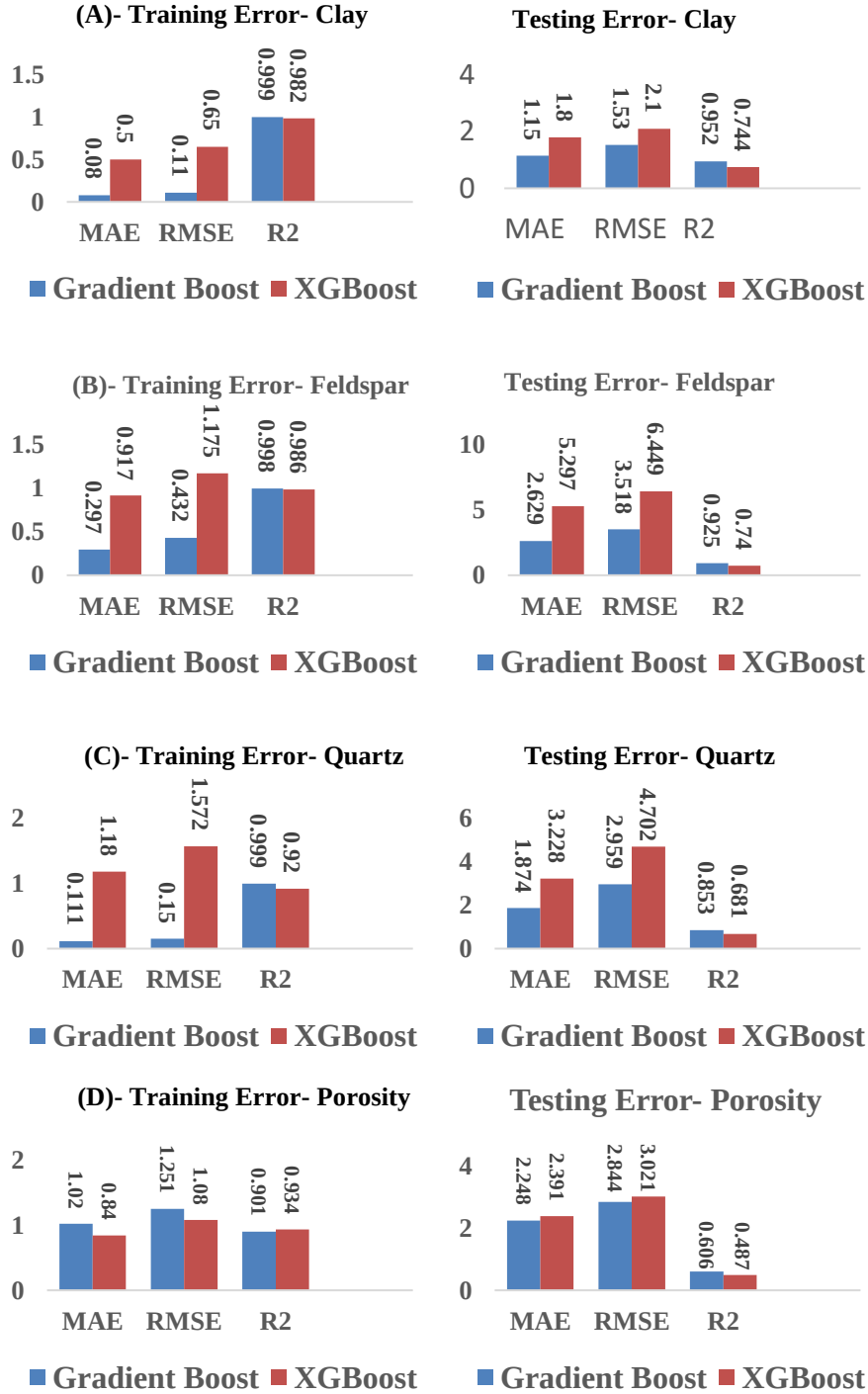


Fig. 5.27: Estimation error comparison of training and testing results for clay, feldspar, quartz contents, and porosity prediction models.

Appendix IX List of Abbreviations

TOC: Total Organic Carbon
MLR: Multiple Linear Regression
ML: Machine Learning
AI: Artificial Intelligence
MAE: Mean Absolute Error
RMSE: Root Mean Square Error
R²: Coefficient of Determination (Square of the correlation coefficient)
XGBoost: Extreme Gradient Boost
 $\Delta\log R$: Change in Logarithmic Resistivity (used in the Passey method for identifying organic-rich source rocks)
XRD: X-Ray Diffraction
Log R: Logarithmic Resistivity (used in geophysical logging)
SEM: Scanning Electron Microscopy
PE: Photoelectric Log
GR: Gamma Ray log
CN: Compensated Neutron log
PE: Photoelectric log
AC: Acoustic log
DEN: Density log
Rt: True Resistivity
SP: Spontaneous Potential log
PSO: Particle Swarm Optimization
FA: Factor Analysis
AAI: Absolute Acoustic Impedance
RAI: Relative Acoustic Impedance
MLFN: Multi-Layered Feed Forward Neural Network
R: Deep resistivity log (ohm-m)
 Δt (AC): Sonic transit time ($\mu\text{s}/\text{ft}$ or $\mu\text{s}/\text{m}$)
R_{Baseline}: Resistivity value at the baseline
 Δt _{Baseline}: Sonic transit time value at the baseline
LOM: Local Level of Organic Metamorphism (a constant related to thermal maturity)

Appendix Own Results Summary

Scientific Question Solved

This research addressed the challenge of accurately characterizing shale reservoirs by concentrating on estimating and evaluating Total Organic Carbon (TOC), mineral composition, and porosity using well logs. The study sought to bridge the divide between traditional evaluation methods and advanced machine learning techniques, thereby improving prediction accuracy and fostering a more comprehensive understanding of shale reservoirs. A key investigation centered on the effectiveness of integrating multiple predictive models to enhance the assessment of shale reservoirs in Formation A of the Ordos Basin.

Methods Personally Applied

The following methods were applied in this research.

1. Development of Predictive Models

- a. I created and applied predictive frameworks utilizing Passey's $\Delta\log R$ approach, Multiple Linear Regression (MLR), and advanced machine learning methods such as Gradient Boost, Adaptive Gradient Boost, Random Forest, and the Stacking Ensemble method.
- b. Conventional well log data (e.g., gamma ray, resistivity, density, Compensated Neutron log, and neutron logs) and core data from XRD for clay, feldspar, quartz and core porosity were collected and analyzed to establish robust models.

2. Data Integration and Interpretation

- a. I incorporated the model outputs into petrophysical software for detailed visualization of lithological and petrophysical characteristics.
- b. Machine learning predictions were combined with well-log responses to identify high TOC zones and delineate brittle and ductile layers within the shale formation.

3. Model Validation and Analysis

- a. Statistical metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 (coefficient of determination) were applied to assess and compare model performance.
- b. Sensitivity analyses were also conducted to evaluate the influence of input variables on predictive accuracy.

Results Obtained

1. TOC Evaluation

- a. The TOC successfully demonstrated strong correlations between predicted TOC values and core-measured data using Passey's method, MLR, and machine learning techniques.
- b. In this research, the Gradient Boost models and Stacking technique showed superior performance in the training and testing phases and, as such, outperformed other methods, achieving higher prediction accuracy with minimal errors, particularly in organic-rich zones.

2. Mineral Composition and Porosity Prediction

- a. This work successfully predicted the mineralogical composition (clay, feldspar, and quartz) and porosity of Formation A shale using multiple linear regression and machine learning techniques.
- b. Gradient Boost models exhibited exceptional training and testing phase results, providing reliable predictions for reservoir characterization.

3. Integration and Workflow Optimization

- a. Integrating the predictive outputs into petrophysical software to create depth-matched wireline log interpretations, facilitating more precise reservoir characterization.
- b. There is a reduced dependency on traditional core-based methods, offering cost-effective and scalable alternatives through data-driven techniques.

4. Practical Contributions to Reservoir Management

- a. This research identified zones with high TOC and favorable brittleness characteristics, which aid in selecting optimal hydraulic fracturing targets.
- b. There is an improved understanding of heterogeneity in shale reservoirs, leading to better-informed exploration and production strategies.

In conclusion, this research demonstrates the transformative potential of integrating traditional geophysical methods with machine learning models, providing a framework for enhanced evaluation and management of unconventional reservoirs globally. This research tends to improve the already traditional method, which is time-consuming and expensive. The reservoir is usually not wholly informed due to the lack of continuous samples within the reservoir of interest using the traditional method but wholly informed by the Machine learning methods used in this research.