Journal of Himalayan Earth Sciences Volume, No., 2026 pp. 1-25

Electrofacies and Petrophysical Characterization of the Datta Formation in the Kohat Plateau, Northwest Pakistan

Syed Asad Ali Kazmi*[1,] and Dr. Nosheen Akhtar[1]

[1] *Institute of Geology, University of the Punjab, Lahore, Pakistan*

**Corresponding Author: asad51472@gmail.com*

Submitted Date:12/12/2025, Acceptance Date: 04/05/2026, Publication Date:---------

Abstract

Understanding reservoir heterogeneity within the Datta Formation remains a key challenge for hydrocarbon exploration and development in the Kohat Plateau, northwestern Pakistan, due to laterally variable sandstone bodies interbedded with shale and limited subsurface control. This study aims to improve reservoir characterization by integrating petrophysical analysis with electrofacies classification to better resolve lithological variability and reservoir quality distribution better.

The study focuses on the Jurassic Datta Formation across the Kohat Plateau and utilizes wireline log data from seven wells distributed throughout the basin. The dataset includes gamma ray (GR), density (DEN), neutron porosity (NPHI), compressional slowness (DTC), and deep resistivity (Deep Res) logs, which collectively capture lithological and fluid-related responses within the formation.

We applied an integrated workflow combining conventional petrophysical evaluation with an unsupervised machine-learning approach. Principal component analysis followed by k-means clustering was implemented using the Heterogeneous Rock Analysis (HRA) module in Techlog to classify electrofacies without predefined lithological constraints.

The analysis identifies six distinct electrofacies ranging from clean, high-quality sandstone reservoirs to tight, non-reservoir siltstone and shale facies. The electrofacies classification was validated using a confusion matrix and accuracy metrics, resulting in an overall model accuracy of 79.5%, indicating reliable discrimination between reservoir and non-reservoir facies. These facies distributions are consistent with observed reservoir behavior across the basin. This study represents the first basin-scale integration of petrophysical analysis with HRA clustering for electrofacies characterization of the Datta Formation in the Kohat Plateau, providing a robust framework for improved reservoir prediction and development planning.

Keywords: Kohat Plateau; Heterogeneous Rock Analysis (HRA); Principal Component Analysis (PCA); Electrofacies, Petrophysics.

1. Introduction

Petrophysical evaluation is a fundamental approach for reservoir characterization in clastic sedimentary successions, particularly in settings where continuous core control is limited. These analyses provide quantitative estimates of porosity, shale volume, water saturation, and log-derived permeability, forming the basis for lithological and reservoir quality interpretation (Asquith and Krygowski, 2004; Ellis and Singer, 2007). In the Datta Formation of the Kohat Basin, the presence of vertically stacked and laterally variable sandstones interbedded with siltstone and shale units reflects a high degree of depositional heterogeneity, which complicates lithological discrimination and reservoir assessment when conventional petrophysical techniques are applied in isolation (Miall, 2010). Subsequently, this study investigates the integration of conventional petrophysical analysis with the Heterogeneous Rock Analysis (HRA) workflow to establish a robust electrofacies framework and improve reservoir characterization of the Datta Formation.

For effective prospect evaluation and reservoir characterization, proper facies identification is essential. Previously, the facies were determined by traditional methods, e.g., cross-plots and their integration with core descriptions. In recent years, many machine learning methods based on mathematical computations have been conceived for facies identification. This allows geoscientists to develop a better understanding of rock geometry with greater precision and understanding of reservoir behavior on larger scales in field development plans.

The reservoir rock types control the distribution and quality of petrophysical characteristics, which is significant for reservoir characterization (Cherana et al., 2022; Doghmane et al., 2022). It enhances the knowledge of pore size distribution, permeability, lateral and vertical heterogeneity, and storage capacity of the rock. For rock type classifications, many different criteria may be used.

In rock type classifications, the reservoir rocks are divided into distinct units that were shaped under equivalent geological settings and have endured similar diagenetic changes (Guo et al., 2005). Petrophysical properties in a reservoir rock are the fundamental parameters that control the quality of reservoir rock (Mahmood and Guo, 2019) and need to be considered in reservoir rock typing. This research combines rock classification with knowledge of petrophysical parameters. The term electrofacies refers to the set of log responses that identify a single stratigraphic unit 'bed' and differentiate it from the others, introduced by Serra and Abbott (1982) and became an essential part of reservoir characterization methodologies (Kumar and Kishore, 2006). The log measures the physical properties of the rocks; electrofacies may specifically be attributed to one or more lithofacies. By clustering log responses based on their inherent relationships, this approach reveals natural facies trends without prior lithological assumptions (Everitt et al., 2011; Jain et al., 1999).

Electrofacies and multivariate statistical techniques have been widely applied for reservoir characterization, permeability prediction, and rock typing in both clastic and carbonate systems, demonstrating their ability to capture lithological heterogeneity beyond conventional log interpretation (Lee et al., 2002; Radhy and Nasser, 2017; Davis, 2018; Rojas et al., 2020). These approaches exploit the sensitivity of wireline log responses to variations in lithology, texture, and depositional attributes and are commonly integrated with established petrophysical methods for estimating key reservoir properties (Rider, 1996; Schlumberger, 1989). Despite the proven effectiveness of such techniques, no basin-scale electrofacies study has previously been conducted for the Jurassic Datta Formation in the Kohat Plateau, where reservoir quality is strongly influenced by lateral facies variability and compressional tectonics. This study addresses that gap by integrating conventional petrophysical analysis with Heterogeneous Rock Analysis (HRA) clustering to establish a regionally consistent electrofacies framework, providing new insights into the distribution of reservoir and non-reservoir facies and supporting improved exploration and development strategies in the Kohat Basin.

2. Study Area

A regional tectonic framework of Pakistan (Fig. 1), showing plate boundaries, fold-and-thrust belts and major structural elements, follows the conceptual tectonic model of the northern Indian plate–Eurasian plate collision as illustrated by Beck et al. (1996).

The study area, Kohat Plateau (Fig. 2), is situated in the northwestern Himalayan foreland fold and thrust belt of Pakistan. This thrust belt was developed because of the continuing collision between the Indian and Eurasian plates (Kazmi and Jan, 1997). The basin is structurally characterized by thrust faulting and tight folding, impacting the variability in thickness and reservoir properties (Bilal et al., 2023). The plateau is illustrated by an asymmetric fold-and-thrust geometry, with broad synclines, numerous tight anticlines, and substantial thrust faults that accommodate crustal shortening in the region (Yeats and Thakur, 2008).

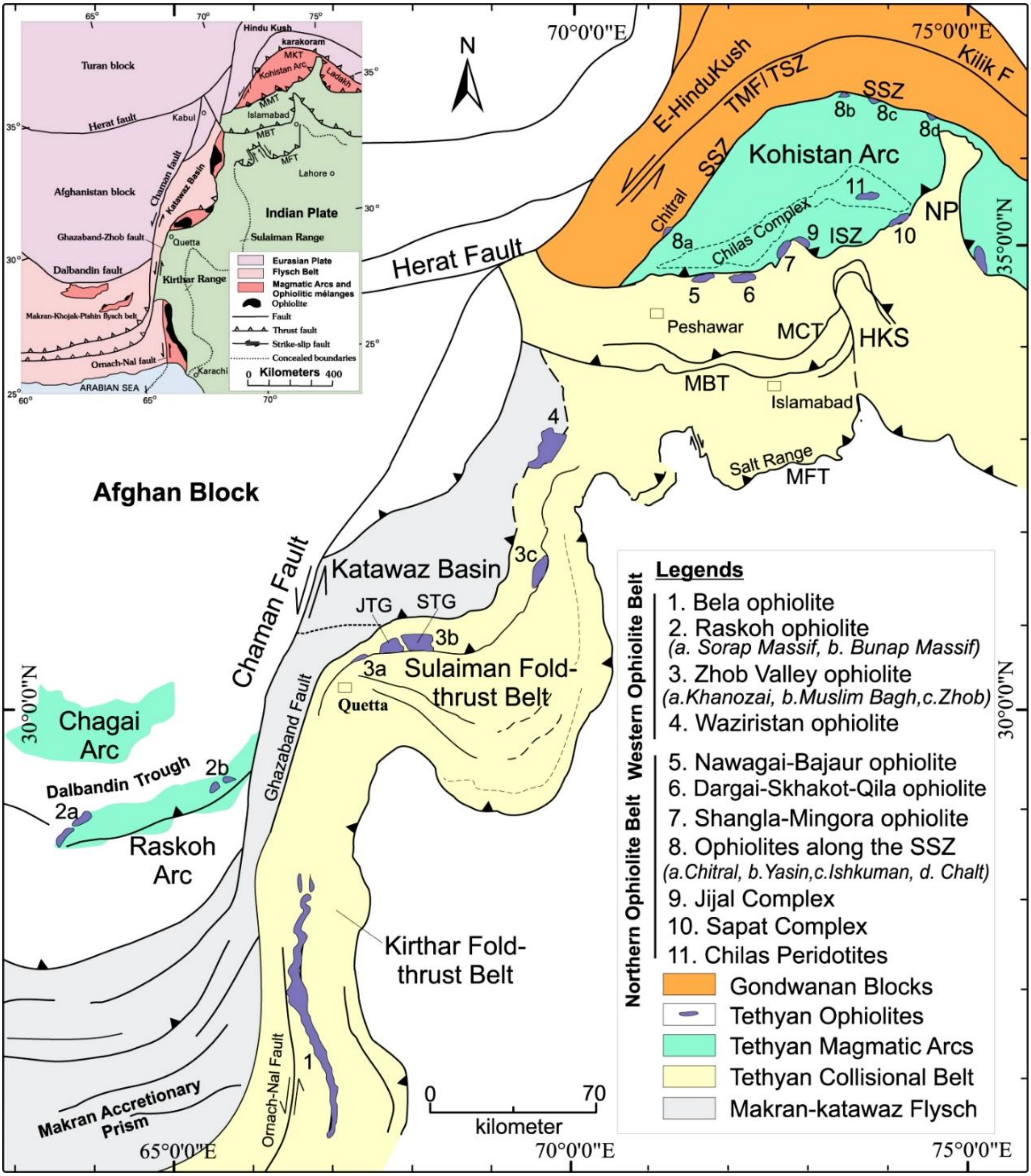

Fig. 1. Tectonic map of Pakistan after Beck et al. (1996).

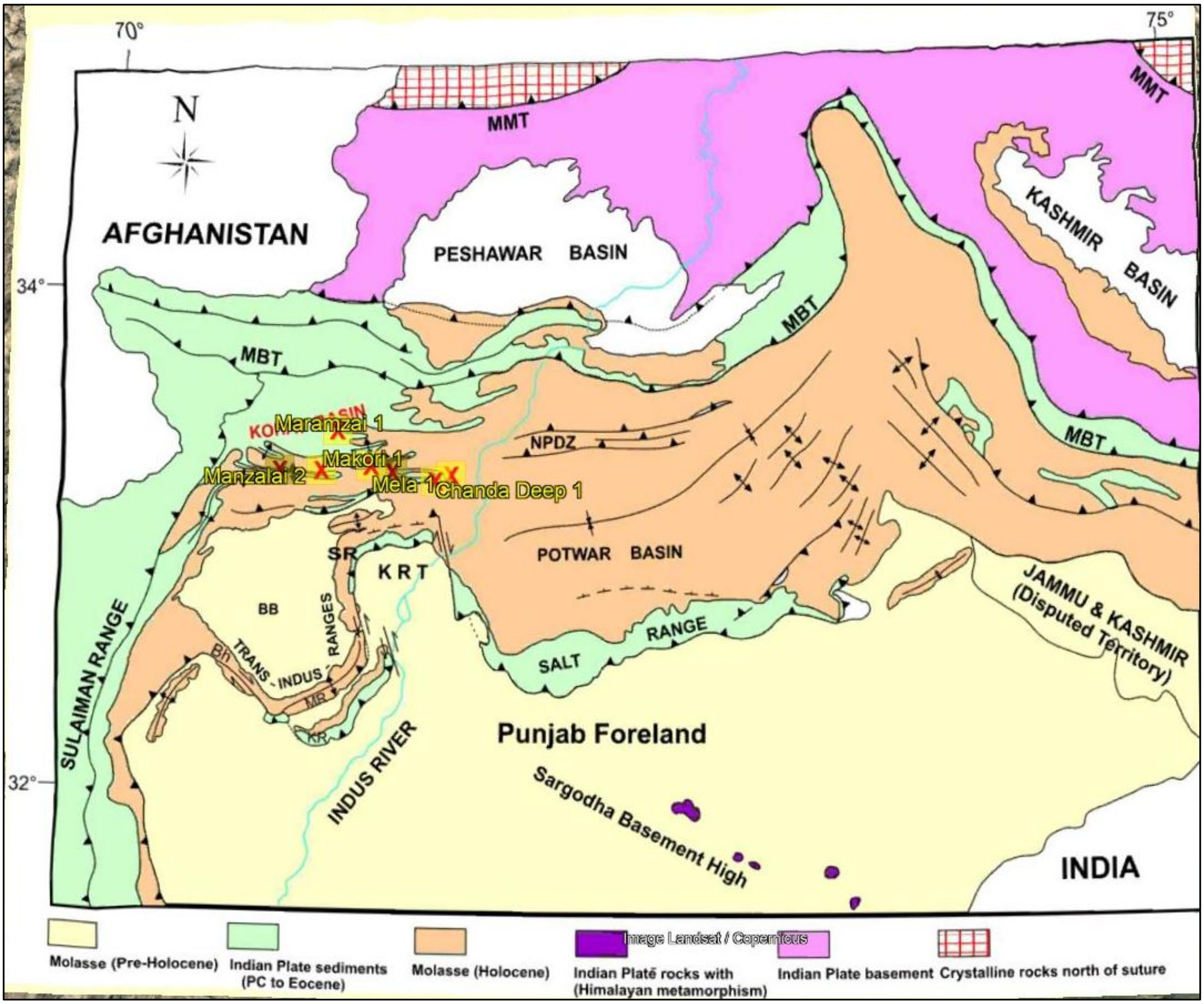

Fig. 2. Tectonic map of Kohat Plateau after Meissner et al. (1974); Pivnik and Sercombe (1993).

A regional tectonic framework of the Kohat Plateau, showing major structural boundaries, fold–thrust geometry, and fault trends, has been widely presented in geological maps and cross sections modified after Meissner et al. (1974) and Pivnik and Sercombe (1993).

The Kohat Plateau comprises a thick stratigraphic succession of Mesozoic to Cenozoic sedimentary rocks, which includes Jurassic to Cretaceous siliciclastic and carbonate sequences. These are overlaid by Paleogene and Neogene molasse deposits. These formations were deposited in shallow-marine to continental environments, which later deformed during the Himalayan orogeny (Meissner et al., 1974; Pivnik and Sercombe, 1993). Numerous important hydrocarbon-bearing units are hosted in the Kohat plateau, where potential and proven reservoir targets are commonly found within the Datta Formation, Lumshiwal Formation**,** Lockhart Limestone, and Paleogene clastics. The development of structural traps, thrust stacking, and tectonic uplift have played a vital role in shaping the petroleum system of the region (Fatmi, 1973; Dani et al., 1995).

Overall, the Kohat Plateau represents a dynamic tectono-sedimentary setting where foreland deformation, stratigraphic complexity, and active structural growth significantly influence reservoir distribution and quality.

In general, the Kohat Plateau is a very significant part of the Upper Indus Basin with respect to hydrocarbon exploration. Different oil and gas fields are present here, e.g., Nashpa, Mela, Chanda, Mamikheil, Mamikheil South and East, Togh, Sahib Gul.

3. Geology of Datta Formation

The Datta Formation is a very significant part of hydrocarbon exploration in the Kohat Plateau, keeping in view the stratigraphic sequence and hydrocarbon exploration history. The reservoir sands of the Datta Formation have been discovered for oil and gas over the last two decades. The Datta Formation is composed of fluvio-deltaic sandstone and shale strata (Fig. 3) from Khan and Rehman (2019). The sandstones are of continental derivation. Generally, very clean, mostly quartzose, fine- to coarse-grained, friable, fine to coarse grained and slightly carbonaceous, having a very high range of porosity (Kadri, 1995). The individual beds are up to 25 meters thick. Features of weathering, variable colors mostly red, maroon, grey, and white, are visible in Datta sands; that's why the variegated series name was given by previous experts. Every single bed of the Datta Formation characterizes a fine upward cycle. Intercalated and laminated sandy shales having some carbonaceous content are also observed with the sandstones. The formation has a versatile depositional environment, which represents distributary channel, overbank and abandoned channel, swamp, interdistributary bay, delta front and mud flat. In general, the Datta sands have excellent petrophysical properties. (Khan and Rehman, 2019).

4. Methodology

For this study, wireline log data from seven wells have been used, which were drilled across the Kohat Plateau penetrating the Datta Formation. The wells include Chanda Deep-1, Chanda-1, Mela-1, Well X-1, Makori-1, Manzalai-2, and Maramzai-2. Three wells (Makori-1, Manzalai-2, and Maramzai-1) have core and cuttings flags, which were used to calibrate the electrofacies model. These wells have a spatial distribution from north to south of the Kohat plateau. The wireline datasets consist of Gamma Ray (GR), Density (DEN), Neutron Porosity (NPHI), Compressional Slowness (DTC), and Deep Resistivity (Deep Res) logs. The location map of Kohat Plateau shows the distribution of the studied wells (Fig. 2).

Petrophysical analysis was integrated with multivariate HRA-based clustering in Techlog to derive electrofacies and heterogeneity trends across the Datta Formation, as illustrated in Figure 4.

Quantitative insights into key rock properties such as porosity, permeability, water saturation, and lithology are obtained by petrophysical analysis, which plays a vital role in reservoir characterization. Integration of well log data with core measurements allows the estimation of reservoir quality, heterogeneity, and fluid distribution, providing the foundation for precise subsurface modeling (Tiab and Donaldson, 2015; Asquith and Krygowski, 2004). Moreover, combining statistical or machine-learning approaches, such as clustering or electrofacies analysis with petrophysical parameters, enhances the identification of discrete reservoir units and improves the understanding of spatial variability within the reservoir (Serra, 1984-1986). Such integrated analyses are essential for optimizing exploration, development, and production strategies in hydrocarbon reservoirs.

Heterogeneous Rock Analysis (HRA) clustering technique was used to generate rock types by utilizing multivariate input data. The input data is converted to separate axes by using Principal Component Analysis (PCA) in the module, front- loading the variance and making it certain that the data used in clustering is functionally self-determining. It permits us to filter out the noise from the input data and to see the variation in the data more clearly. In the first step, the model used the heterogeneous rock analysis (HRA) clustering procedure to classify rock units. Subsequently, principal component analysis (PCA) evaluates the data from the procedure in the first phase to make sure that they are independent (Hotelling, 1933; Pearson, 1901). The multidimensional geometric rotation takes place in this process, which brings into line with the orthogonal axes of maximum disparity within a data point cloud (Doveton, 1994; Kafayatullah et al., 2006).

The performance of the model was evaluated by using a confusion matrix and standard accuracy metrics. In this process, the comparison was made between predicted electrofacies and core- and cutting-based facies flags. The overall accuracy of the computed classification was 79.5%, which indicates strong and geologically reasonable facies discrimination.

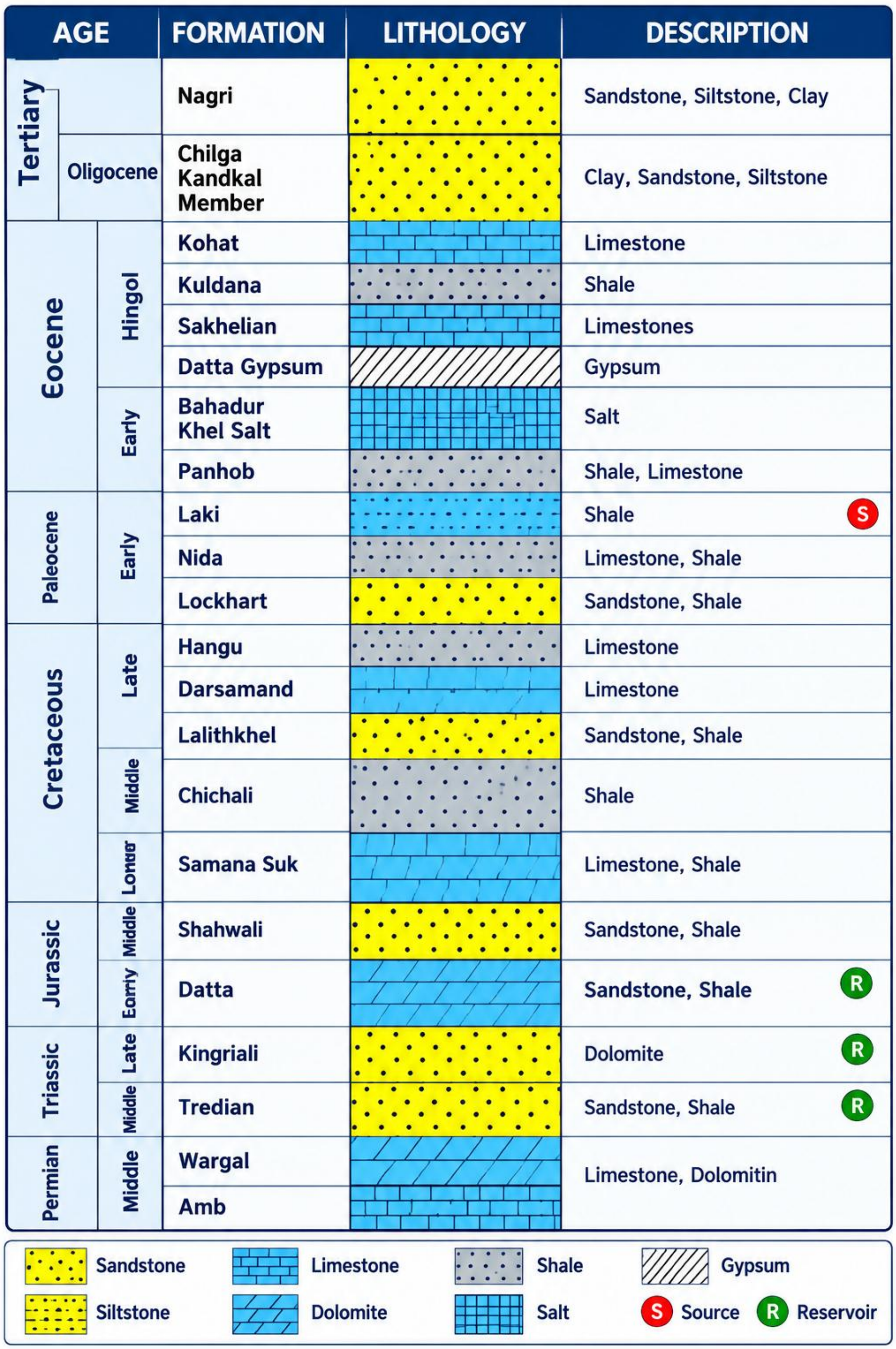

Fig. 3. Generalized Stratigraphy of Kohat Basin after Khan and Rehman (2019).

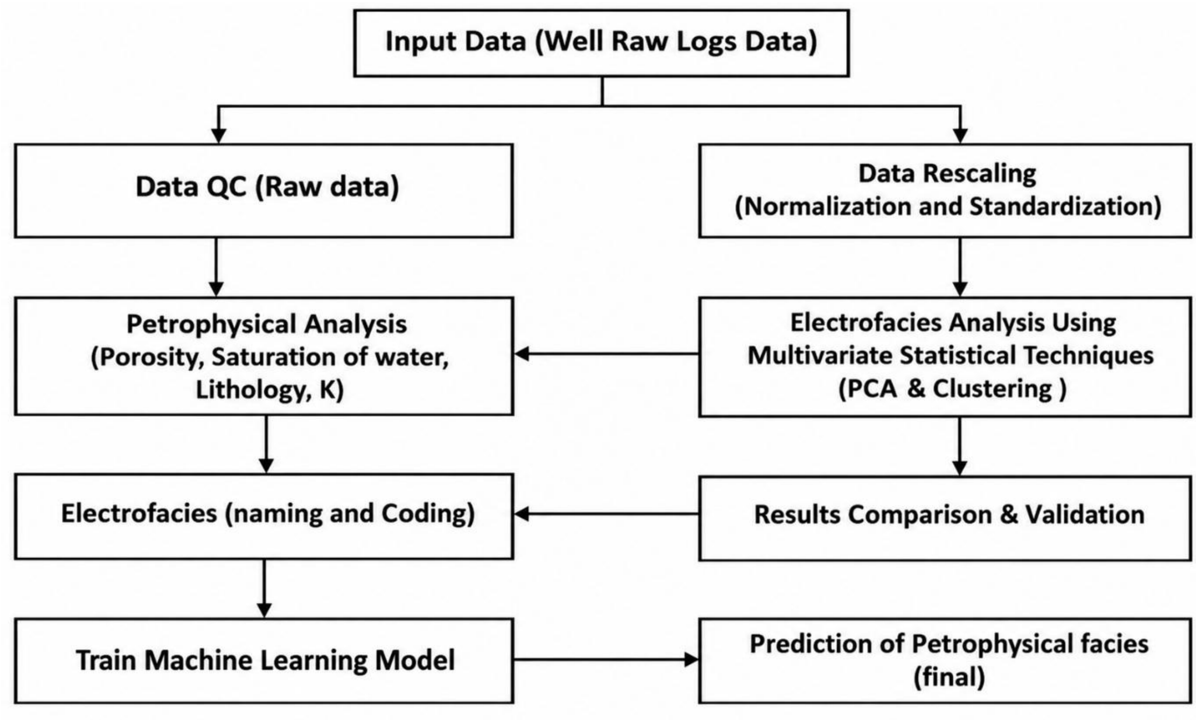

Fig. 4. Integrated workflow for the study.

5. Results and Discussion

5.1. Petrophysical Analysis for Reservoir Characterization

Petrophysical evaluation of the Datta Formation was undertaken to assess reservoir quality, delineate lithology, and hydrocarbon potential across the Kohat Plateau

5.1.1 Data visualization and QC

During the process of creating the petrophysical database, considerable effort was made to prepare the initial data. All available log, survey, and core data were imported and quality-checked in preparation for evaluation. Logs were spliced as needed to create single-instance curves over the entire well.

Adjustments were made to the well logs, including environmental corrections and depth-shifting. Gamma ray logs were in good quality overall. In some wells at depth with washouts, density logs, neutron logs, and sonic logs were affected due to reading of mud fluid instead of formation, but overall, the data quality was good in entre logged intervals. All log curves were compared using histogram frequency analysis to assess anomalous logs and other log quality issues.

5.1.2 Lithology Computation

A probabilistic well-log analysis methodology was used to estimate the lithology in this study. Multi-mineral modeling involves the simultaneous solution of multiple linear and non-linear equations. Inputs include logging data, formation water salinity, formation temperature, and mud properties. Model results include the volume fractions of specified formation minerals and fluids. Multi-mineral modeling integrated these advanced logs and reduced the uncertainty of the petrophysical model. Main minerals in the model included illite and quartz. Due to high formation water salinity and alternating layers of shale, the Indonesian water equation was chosen to evaluate the fluid saturations to address the shaly section within the reservoir intervals, with a cementation exponent (m) of 2 and a saturation exponent (n) of 2. The electrical property parameters of the saturation equation were from standards that are measured and established empirically in the absence of special core analysis (SCAL). Average salinity of 64000 ppm NaCl equivalent, at average bottom hole temperature 120 DEGC, has been calculated by using the top log interval and bottom log interval to estimate formation water resistivity. The salinity has been taken from the regional formation water salinity values of the Datta Formation in the area. The Indonesian water equation is designed to work in shaly intervals too. Input logs of multi-mineral modeling included basic logs such as GR, RHOB, NPHI, DTC, RT, and shallow resistivity log (RXO). Outputs of the multi-mineral model included mineral volume and weight fractions, total and effective porosity, and water saturation.

5.1.3 Volume of Shale

Before estimation of VSH, the lithology of Datta Formation was reviewed based on multi-mineral modeling. The matrix of the reservoirs consisted of quartz; the main components of clay were illite. Due to the presence of K-feldspar, VSH was estimated by using a combination of GR and NPHI-RHOB cross-plot based on the log availability and the log quality.

The GR log measures natural radioactivity of the rock and is expressed in American Petroleum Institute units (API). The gamma ray index (GRINDEX) was calculated next as a component of the estimation of VSH:

$$GRINDEX = (GR - GR_{SAND}) \; / \; (GR_{SHALE} - GR_{SAND}) \tag{1}$$

VSH_GR was estimated using the Clavier equation (2):

$$VSH\_GR = 1.7 - \sqrt{3.38 - (GRINDEX + 0.7)^2} \tag{2}$$

The sand and shale endpoints were varied with depth to approximate the 10th and 90th percentile of the GR log response as the GR log varied up and down in the well.

An alternative method to estimate VSH is from Neutron and density logs. In clean sandstones, the neutron and density porosity values are similar and often overlap, assuming that each porosity curve is computed on a sandstone matrix. In shale, the porosity curves diverge because of a lower shale matrix density, which causes a reduction in density porosity and higher neutron porosity as a result of the hydrogen index effect in clay. VSH_ND was estimated using the relationship (3):

$$VSH\_ND = \frac{\frac{NPHI_{matrix} - NPHI}{NPHI_{matrix} - NPHI_{fluid}} - \frac{RHOB_{matrix} - RHOB}{RHOB_{matrix} - RHOB_{fluid}}}{\frac{NPHI_{matrix} - NPHI_{shale}}{NPHI_{matrix} - NPHI_{fluid}} - \frac{RHOB_{matrix} - RHOB_{shale}}{RHOB_{matrix} - RHOB_{fluid}}} \tag{3}$$

Where:
NPHIfluid = neutron porosity log of the fluid, decimal
NPHIshale = the neutron porosity log of shale, decimal
NPHImatrix = the neutron porosity log of the matrix, decimal
RHOBmatrix = the bulk density of the matrix, g/cc
RHOBfluid = the bulk density of the fluid, g/cc
RHOBshale = the bulk density of shale, g/cc

Based on the log availability, VSH_GR was estimated in 07 wells. VSH_ND was estimated in 2 wells with both neutron and density logs. Because VSH_GR was overestimated at intervals, the final VSH (VSH_Final) in was estimated using the minimum value of VSH from the two methods.

5.1.4 Porosity Estimation

Three different classes of porosity-oriented logging tools were available for this study, such as DT, RHOB, and NPHI. All the wells have acoustic logs, neutron logs, and density logs. Based on the availability of porosity logs, different types of porosity were estimated for each well.

$$DPHI_{cc} = DPHI - VSH*DPHI_{shale} \tag{4a}$$
$$NPHI_{cc} = NPHI - VSH*NPHI_{shale} \tag{4b}$$

Where, $DPHI_{cc}$ = density effective porosity after correcting for shale, decimal
DPHI = density total porosity in sandstone units, decimal
$DPHI_{shale}$ = density total porosity in sandstone units for shale, decimal
$NPHI_{CC}$ = neutron effective porosity after correcting for shale, decimal
$NPHI_{shale}$ = neutron total porosity for shale, decimal

Neutron-Density cross plot is used to compute porosity and delineate lithology (Fig. 5a and Fig. 5b).

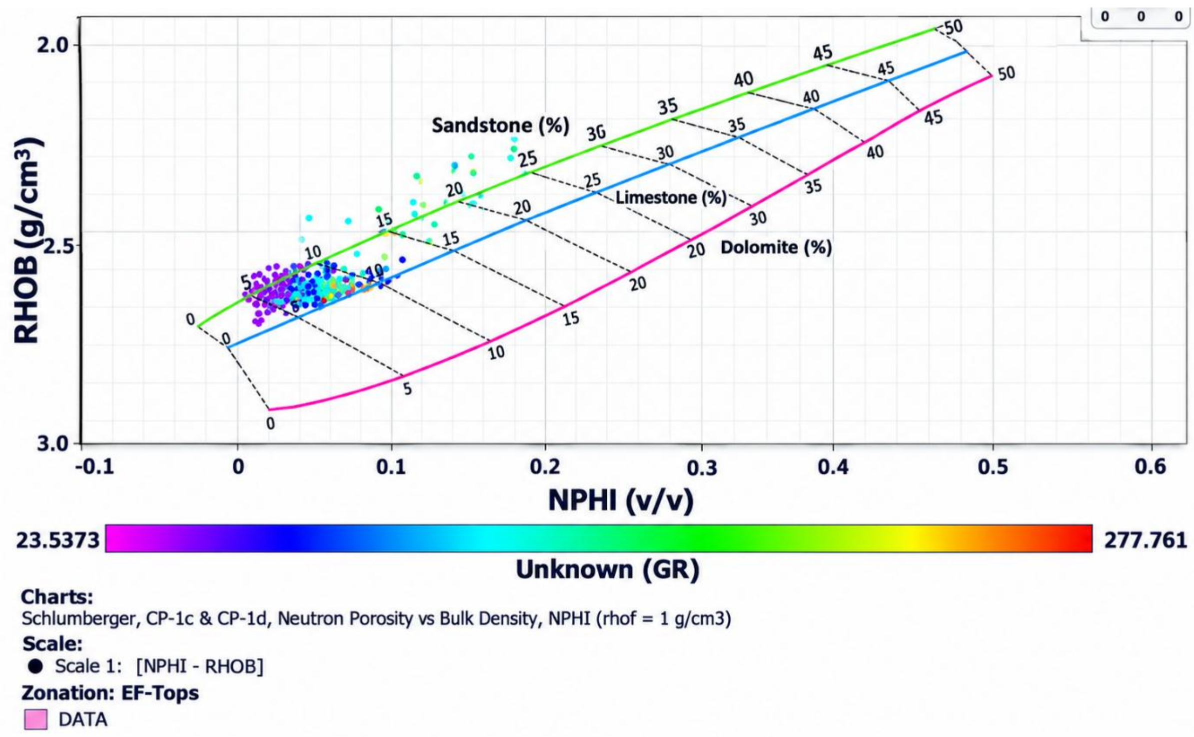

Fig. 5a. N-D Cross plot for lithology delineation for petrophysical analysis of Chanda Deep-1.

5.1.5 Sonic Porosity

The sonic log is available in 07 wells, which was used to estimate PHIE (PHIE_S). The sonic Wyllie time average equation was applied for calculating porosity.

$$SPHI = (\Delta t_{log} - \Delta t_{ma}) \ / \ (\Delta t_{fl} - \Delta t_{ma}) / \ Cp \quad (5a)$$

$$SPHI\_shale = (\Delta t_{shale} - \Delta t_{ma}) \ / \ (\Delta t_{fl} - \Delta t_{ma}) / \ Cp \quad (5b)$$

Where SPHI = total sonic porosity, decimal; SPHI_shale = shale sonic porosity, decimal; Δt_{log} = acoustic transit time log measurement, µsec/ft; tma = matrix transit time calibrated to match core, µsec/ft; tshale = shale transit time, µsec/ft; tfl = fluid transit time, µsec/ft; and Cp= compaction correction, tuned to match core data.

The resulting sonic porosity was then corrected for shale effects to obtain an effective sonic porosity curve:

$$PHIE\_S = SPHI - SPHI\_shale * VSH \quad (6)$$

Where, PHIE_S = effective porosity from sonic logs, decimal.

5.1.6 Density Porosity

Density porosity was estimated in 08 wells with available density logs but not neutron logs. It was corrected for shale effects to obtain an effective neutron porosity curve in these wells:

$$PHIE\_D = DPHI - DPHI_{shale} * VSH \quad (7)$$

Where, PHIE_D = effective porosity from density logs, decimal.

5.1.7 Final Porosity

Final porosity was mainly estimated from a combination of porosities from different methods. In washout zones, PHIE_S was used to estimate final PHIE. Log-calculated porosity was observed to have a very good correspondence

5.1.8 Water Saturation

SWE estimates require PHIE, VSH, a deep reading resistivity log, and knowledge of the formation water resistivity (Rw). Reservoir temperature was estimated by a top log interval and bottom log intervals, using the surface temperature and bottom-hole measured temperature.

Formation water salinity of 64 thousand parts per million (kppm) was used in this study. Due to high formation water resistivity and shaly sand, effective water saturation was estimated using the Indonesian equation of the form:

$$R_o = \frac{a * R_w}{\phi^m}; \qquad C = 1 - (V_{sh} * 0.5)$$

$$A = \frac{Vsh^{c}}{\left(\frac{Rsh}{Rt}\right)^{0.5}};\qquad B = \left(\frac{R_t}{R_o}\right)^{0.5}$$

$$S_w = (A+B)^{\frac{-2}{n}};$$

$$BVWE = SW_e * \varphi_e \qquad (8)$$

Where, a =Tortuosity Constant (1), m= Cementation Factor (2), n= Saturation Exponent (2), RSH= Shale Resistivity (Ohmm), and Rw = Formation Water Resistivity. The outputs are: SWE_INDO = Saturation Indonesian (v/v), SWE_INDO_UNCL = Saturation Indonesia Unclipped (v/v), and BVWE_INDO = Effective Bulk Volume Water (v/v).

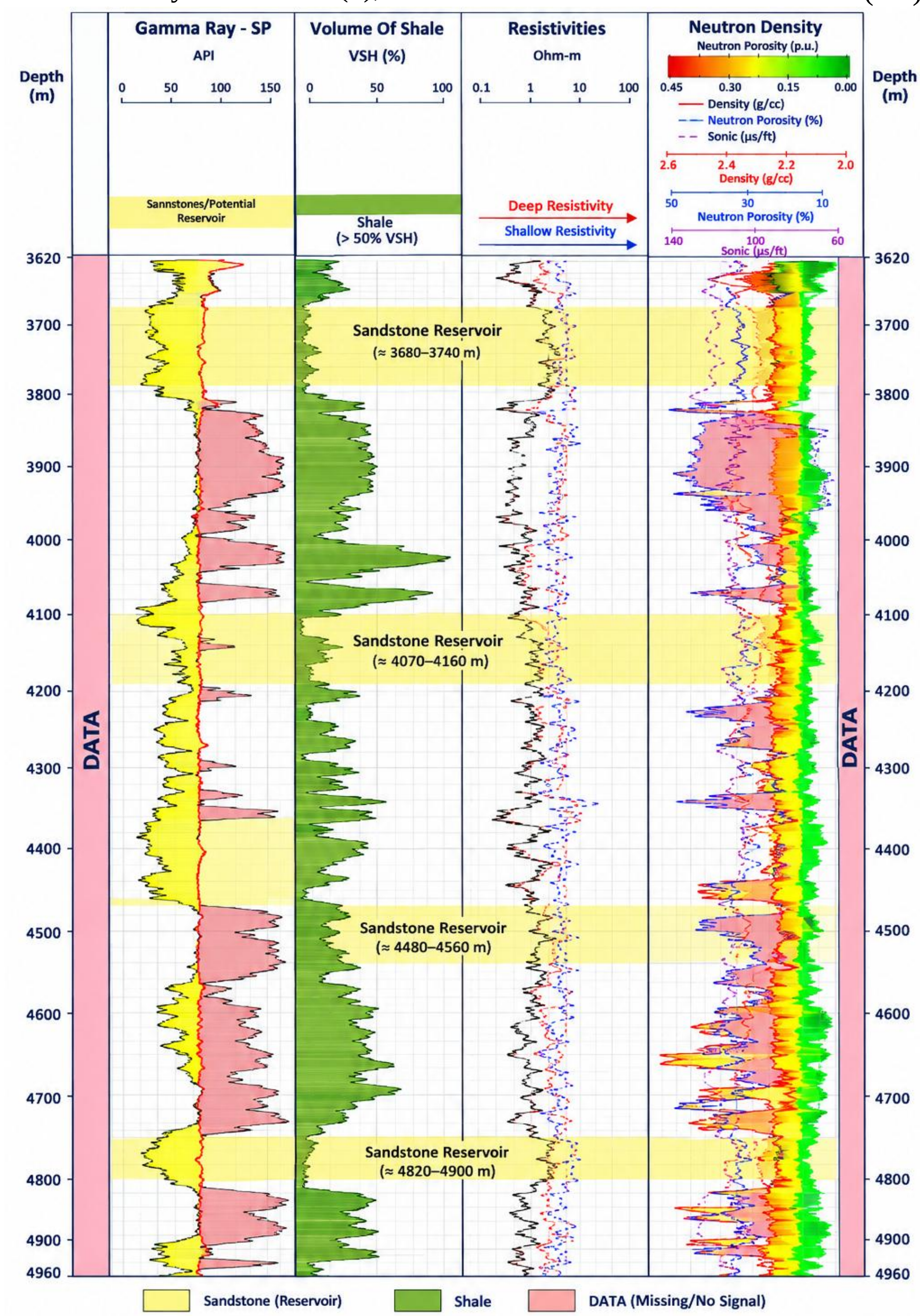

Fig. 5b. Well logs of Data formation in Chanda Deep-1 showing correlations of Gamma ray, Volume of Shale, Resistivity, and Neutron Density logs for petrophysical analysis.

5.1.9 Permeability

Permeability (K) was estimated in 07 wells. The permeability is computed using Coat and Wyllie-Rose methods and utilized as the best fit for final results. Coates permeability is used as the final permeability as the best fit and in correspondence with the computed reservoir characteristics. The wireline log data analysis is used to compute permeability, as core data was not available for the study. Permeability was estimated using the following equations in the Datta Formation. The Coates Equation (9) is:

$$PERM = kc * PHLe^4 * \left(\frac{1 - Swirr}{Swirr}\right)^2 \quad (9a)$$

Else,

$$PERM = kc * PHLe^4 * \left(\frac{PHLt - PHLe * Swirr}{PHLe * Swirr}\right)^2 \quad (9b)$$

The inputs are:

Effective Porosity = Calculated Effective Porosity (v/v), Total Porosity = Calculated Total Porosity (v/v), and Irreducible Water Saturation = Calculated Irreducible Water Saturation (v/v). The output is Permeability = Calculated (mD)

The Wyllie-Rose Equation is:

$$PERM = Kw * \frac{PHI^d}{SW^e} \quad (10)$$

Morris-Biggs d= 6.0 Timur d = 4.4

5.1.10 Net Pay

Net pay flags were estimated based on VSH less than 0.3, PHIE larger than 0.06, and SWE less than 0.6. A sensitivity study has been performed on PHIE cutoffs Table 1.

Table 1: Hydrocarbon summary of the study wells using cutoffs of saturation of water 60%, porosity min. 6%, and volume of shale maximum 35%.

Sr. No.	Well	Flag Name	Top	Bottom	Gross	Net	Net to Gross	Vol_{shale}	Porosity	Sw
1	Chanda Deep-1	ROCK	4646.1	4807.0	160.93	73.43	0.46	0.11	0.05	0.45
		RES	4646.1	4807.0	160.93	26.69	0.17	0.06	0.08	0.36
		PAY	4646.1	4807.0	160.93	25.92	0.16	0.06	0.08	0.35
2	CHANDA-1	ROCK	4751.0	4788.8	37.75	31.38	0.83	0.09	0.05	0.28
		RES	4751.0	4788.8	37.75	14.88	0.39	0.07	0.07	0.20
		PAY	4751.0	4788.8	37.75	14.88	0.39	0.07	0.07	0.20
3	Chanda-2	ROCK	4882.0	4897.1	15.07	7.32	0.49	0.18	0.17	0.24
		RES	4882.0	4897.1	15.07	6.71	0.45	0.17	0.18	0.23
		PAY	4882.0	4897.1	15.07	3.35	0.22	0.20	0.26	0.10
4	Well X-1	ROCK	4250.0	4866.1	616.15	253.14	0.41	0.14	0.04	0.47
		RES	4250.0	4866.1	616.15	45.26	0.07	0.16	0.10	0.25
		PAY	4250.0	4866.1	616.15	43.43	0.07	0.16	0.11	0.24
5	Makori-1	ROCK	3582.9	4310.0	727.10	373.53	0.51	0.11	0.01	0.84
		RES	3582.9	4310.0	727.10	6.40	0.01	0.10	0.08	0.56
		PAY	3582.9	4310.0	727.10	3.05	0.00	0.13	0.10	0.34
6	MANZALAI-2	ROCK	4020.0	4203.0	183.03	49.68	0.27	0.21	0.04	0.38
		RES	4020.0	4203.0	183.03	11.74	0.06	0.17	0.10	0.25
		PAY	4020.0	4203.0	183.03	11.74	0.06	0.17	0.10	0.25
7	MARAMZAI-1	ROCK	3267.0	3362.0	95.00	40.40	0.43	0.11	0.00	0.81
		RES	3267.0	3362.0	95.00	0.00	0.00			
		PAY	3267.0	3362.0	95.00	0.00	0.00			
8	Mela-1	ROCK	4914.0	4951.0	37.03	8.38	0.23	0.19	0.06	0.57
		RES	4914.0	4951.0	37.03	3.20	0.09	0.11	0.10	0.35
		PAY	4914.0	4951.0	37.03	2.44	0.07	0.10	0.10	0.23

Multi-mineral modeling in Chanda Deep-1 demonstrates that the Datta Formation is largely composed of clean sandstone with localized shale streaks, particularly toward the upper section where quartz–clay mixtures dominate (Figs. A1–A2). The CPI workflow delivers a comprehensive set of reservoir properties—including shale volume, lithology, porosity, permeability, and hydrocarbon cut-offs—forming the basis for net-pay computations in all wells. A regional correlation of these parameters across the studied interval is provided in Fig. A2.

5.2 Reservoir Quality in PAY Intervals of the Datta Formation, Kohat Plateau

The hydrocarbon summary table reveals significant heterogeneity in reservoir quality across the studied wells of the Datta Formation in the Kohat Plateau. Evaluation of the PAY zones shows that the formation exhibits generally low to moderate shale content, low to moderate effective porosity, and variable but often favorable water saturation, consistent with a tight clastic reservoir system previously documented in the region (Kadri, 1995; Wandrey et al., 2004).

5.2.1 Average Volume of Shale

The Vsh values in the PAY intervals range from 0.058 to 0.204 v/v, indicating low shale content in wells such as Chanda Deep-1 (0.058) and Mela-1 (0.095), reflecting relatively cleaner sand intervals favorable for reservoir development. Moderate shale content in wells like Chanda-2 (0.204) and Manzalai-2 (0.169), suggesting mixed lithologies where argillaceous components may influence permeability.

The depositional variability within the Datta Formation is reflected by the values of volume of shale, which is illustrated as a system of mixed fluvial-deltaic to shallow marine, where quality of sandstone can vary significantly laterally (Iqbal et al., 2015; Wandrey et al., 2004).

5.2.2 Average Effective Porosity (Φe)

A very low to moderate range of effective porosity is observed in the pay zones, i.e., from 0.003 to 0.255 v/v.
In the well Chanda-2, moderate porosity is computed, which is 0.255 v/v, whereas in Chanda Deep-1 and Chanda-1 the porosities are 0.072 to 0.077 v/v, which shows better development of pore systems and is favorable for hydrocarbon storage.

Maramzai-1 well shows very low porosity, i.e., 0.003 v/v, which exhibits both diagenetic cementation and tight lithologies, and it is typical for lower Jurassic clastics exposed to burial-associated compaction and silica/calcite cementation (Ahmad et al., 1999). The porosity variations in the Datta Formation show its tightness and support the earlier research findings that it is a tight reservoir and presence of a fracture system is essential for good commercial productivity (Iqbal et al., 2015).

5.2.3 Average Water Saturation (Sw)

A wide range of water saturation is observed in the analysis, ranging from 0.098 to 0.807 v/v. The wells Chanda-2, Manzalai-2, and Mela-1 0.098, 0.25, and 0.227 v/v show less volume of water saturation respectively which is <0.25 v/v, and indicate higher hydrocarbon saturation and effective reservoir behavior.

Mixed fluid occupancy is indicated by a moderate Sw range, i.e., from ~0.33 to 0.36 v/v in wells like Chanda Deep-1 and Makori-1. Maramzai-1 shows high Sw greater than 0.8 v/v and resulted in water-bearing non-reservoir sands.

Such variability highlights the influence of local structural traps, diagenetic modification, and facies changes along the Kohat Basin's thrust-controlled geological framework (Ghani et al., 2015).

5.2.4 Overall Behavior of the Datta Formation in the Kohat Plateau

The PAY-zone analysis demonstrates that the Datta Formation behaves as a heterogeneous and compartmentalized reservoir characterized by:

- Variable clean-to-shaly sand ratios
- Low to moderate porosity
- Generally favorable hydrocarbon saturation in several wells

This is consistent with a tight-sand play, where reservoir quality is strongly lithology-dependent and enhanced by natural fracturing associated with Himalayan tectonics, as widely reported for the Kohat Plateau (Kazmi and Jan, 1997; Iqbal et al., 2015).

5.3 Heterogeneous Rock Analysis (HRA)

Electrofacies were computed by using a multivariate workflow in which Principal Component Analysis (PCA) was used for data reduction, followed by k-means clustering within the Heterogeneous Rock Analysis (HRA) framework.

A scatterplot matrix has been used in order to visualize our data and provide its graphical representation. By using visuals such as maps, charts, and graphs, data visualization tools provide comprehensive control to identify the trends, patterns, and outliers within the datasets. It allows better understanding of the relationship among the variables used for the analysis. In this case, the features taken into consideration are seven well logs, which are Density log (DEN), Compressional Slowness (DTC), Gamma Ray log (GR), Deep Resistivity log (Deep Res), and Neutron Porosity log (NPHI).

The variability of the input logs is presented for each well by the histograms. The first feature, bulk density, shows a common trend for the seven wells, with an average value ranging between 2.5 and 2.8. g/cc. Moreover, the neutron porosity features range between 0.02 and 0.38 v/v, and compressional slowness shows a more homogeneous distribution of data around a value of i.e. ~55 to ~68 us/ft across the seven wells. The NPHI feature shows a moderately elongated trend with values turning around 0.4 v/v. The Gamma Ray feature comprises values that range between 10 API and 200 API, and Deep Resistivity ranges from 10 to 1000 ohm-m (Fig. 6 (a-e)).

HRA classifications are generated to find uniqueness based on a similar set of measurements by running the cluster analysis. Breakdown data into k clusters by minimizing the sum of squared distances between each reflection and the centroid of its allocated cluster; by this means, producing compact and well-separated meaningful groups is the objective of the k-means algorithm (MacQueen, 1967; Jain et al., 1999). We have used multidimensional wireline log data for this study.

Well log data were used to delineate six distinct lithological clusters that show variation in properties of rock by applying the HRA cluster analysis. The first two principal components are displayed on the 2D cross-plot for the selected cluster, with the grey stars in the center of each cluster (Figs. 7, 8). The PCA cross-plot displays the distribution of the log-derived dataset in the multidimensional principal-component space, reduced to its first two components for visualization. These two components capture most of the variance present in the original variables—consistent with the objective of principal component analysis to maximize explained variability while minimizing dimensional redundancy (Jolliffe, 2002; Abdi and Williams, 2010). K-means classification has produced six clusters which are evidently distinguished in the transformed space, and each facies class is characterized by a distinct color. The mean position of every cluster is marked by the centroid, which is plotted as a grey star, and following the classical k-means partitioning rule, all the data points are allocated to the nearest centroid (MacQueen, 1967; Hartigan & Wong, 1979).

The clear spatial separation among clusters in the PC1–PC2 space indicates strong multivariate contrast between lithological groups within the Datta Formation. Clusters exhibit characteristic spreads reflecting internal variability, which is typical in heterogeneous mixed siliciclastic shale successions in foreland basin settings (Catuneanu, 2019; Posamentier and Walker, 2006). The elongated patterns of some clusters additionally suggest correlation structures between the input logs, particularly between density-neutron pairs and gamma-ray responses consistent with petrophysical controls on lithofacies (Doveton, 1994). Overall, the PCA cross-plot demonstrates that the six-facies model preserves geologically meaningful separation in the multivariate domain, supporting its validity for electrofacies classification in the Datta Formation across the Kohat Plateau.

In principal component space, the graphic represents a multivariate cross plot of logs, where most of the variability is captured between the logs rather than the original log curves. In this presentation, the dispersion of the data along each axis is related to the degree of correlation between each input log curve. The color coding in (Fig. 8) corresponds to the rock classes that were assigned based on the unsupervised principal components' classification. Every class is characterized by a class centroid marked by a grey star, and individual data points are allocated to the rock class in principal component space whose centroid is the nearest.

The main model is constructed by using the principal components as input for the k-means clustering algorithm following the PCA analysis. K-means clustering aims to minimize the space among each data point and the centroid of its designated cluster and maximize the partition between the centroids of different clusters. The use of diagnostic plots is to see if the number of clusters selected is statistically substantial and has adequate variability.

The map in (Figs. 10, 11), in principal component space the diagnostic plots, indicated by color scheme with the cluster assignments, the change in the cumulative distance for each run of the model is presented in a fall off plot (Fig. 10), and an indication of cluster uniqueness by depicting how many points of a given cluster may have been allocated to an adjoining class in a silhouette plot (Fig. 9) (Willis, 2013; MacQueen, 1967; Kaufman and Rousseeuw, 1990).

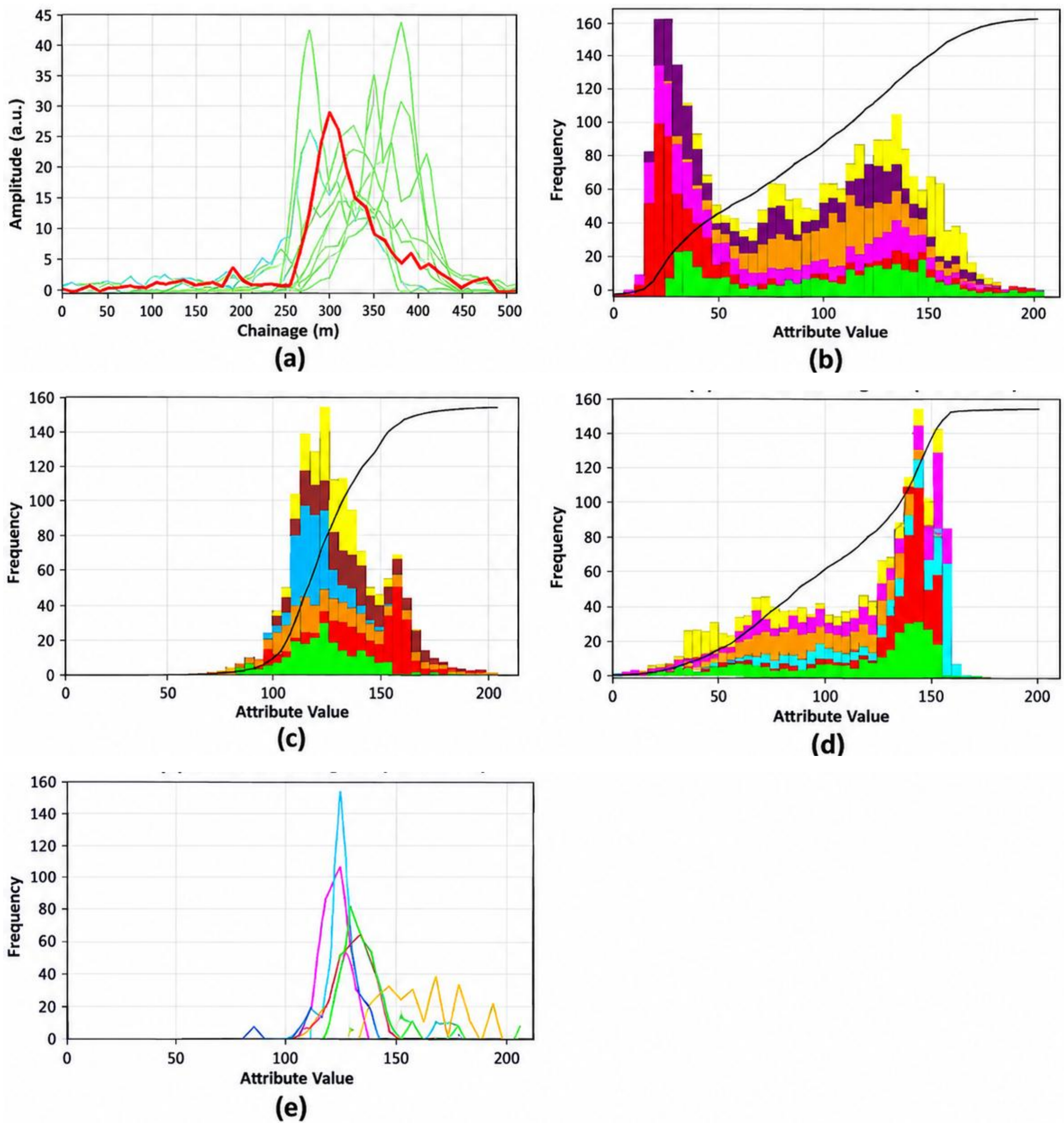

Fig. 6. Histogram of log-transformed variables used in this study.

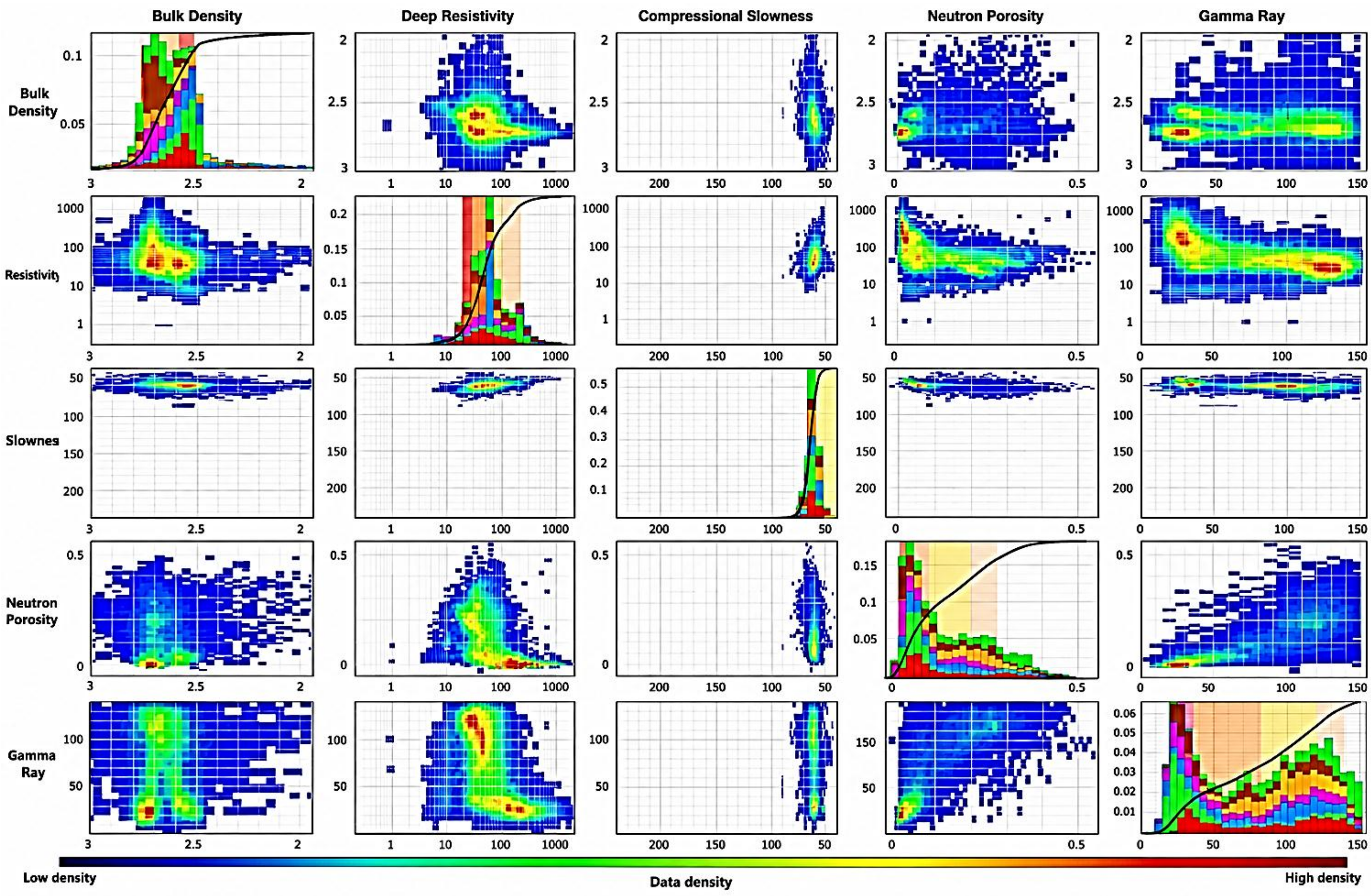

Fig. 7. Matrix Scatter Plot of the features taken into consideration for the HRA method.

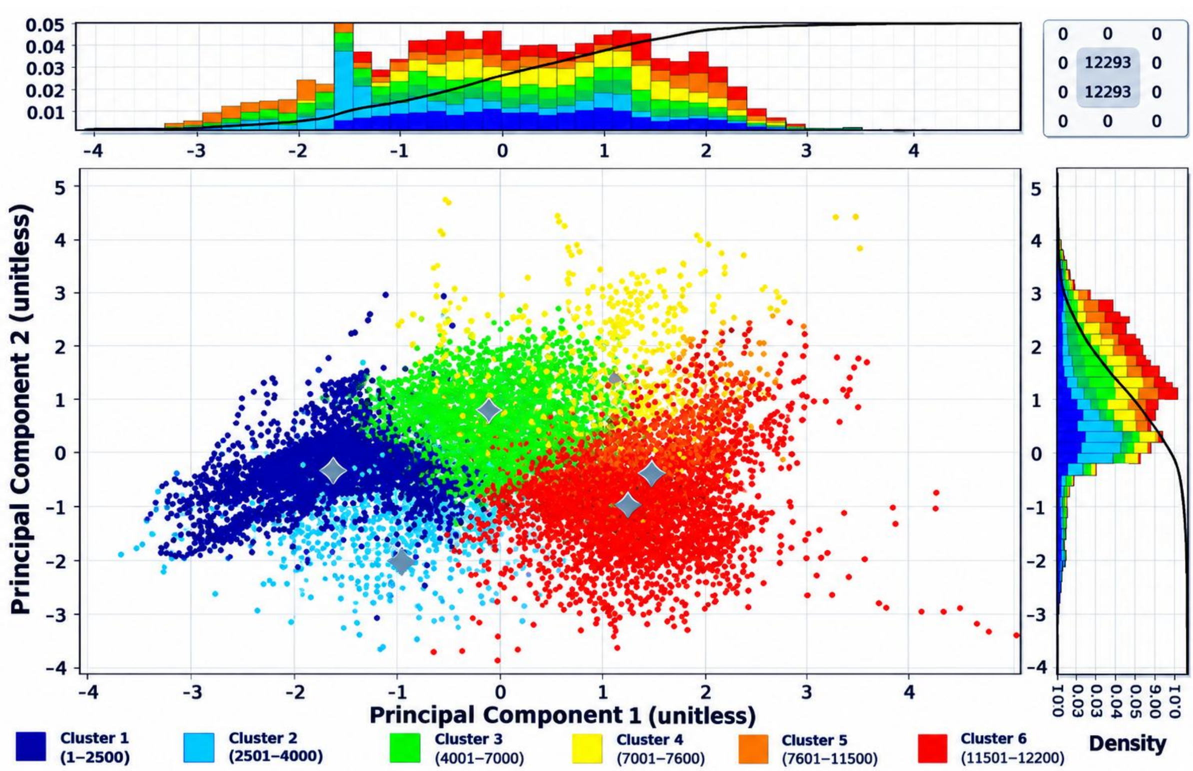

Fig. 8. Cross-plot of the two principal components showing the six clusters of the data.

The HRA silhouette plot provides a quantitative measure of how effectively the six electrofacies clusters represent the multivariate structure of the Datta Formation. Silhouette values evaluate the separation between clusters by comparing the average distance of each point to the centroid of its assigned cluster with the distance to the nearest neighboring cluster (Rousseeuw, 1987). Most of the samples display positive silhouette coefficients in the dataset of Datta Formation, which indicate clusters are well separated in the principal component space, and the detected boundaries of the facies by the algorithm correspond to geologically meaningful transitions.

The PCA cross plot (Fig. 8) demonstrates that, with distinct and coherent fields around their respective centroids, the six clusters reflect the strong inherent heterogeneity in Datta Formation. This heterogeneity is widely documented in previous regional studies that describe the formation as composed of stacked, multi-story sandstone bodies, argillaceous or silty interbeds, heterolithic units, and shale-rich intervals deposited in a continental to marginal-marine setting (Iqbal et al., 2015; Ghani et al., 2015). The clustering pattern therefore aligns with the expected variability in mineralogy and textural maturity recorded in wireline logs.

The clear separation between the clusters suggests that the PCA–k-means workflow is capturing meaningful litho-petrophysical distinctions rather than noise, an outcome consistent with earlier applications of unsupervised clustering in reservoir characterization (Doveton, 1994). Thus, the silhouette results confirm that a six-cluster solution is appropriate for this formation and support the interpretation of discrete electrofacies corresponding to clean sandstones, moderately argillaceous sandstones, siltstones, transitional heterolithics, shale units, and cemented/compacted intervals, each of which is characteristic of the Datta Formation's stratigraphic architecture across the Kohat Plateau.

The cluster analysis fall-off plot shows diminishing cumulative distance with succeeding model runs. A flattening of the cumulative distance within the last 10% of model runs shows that a minimal distance has been reached, and the model is thus adequate.

Six electrofacies have been chosen based on the diagnostic plots and analysis, silhouette plots (Fig. 9), and distance fall-off plots (Fig. 10).

The fall-off plot generated from the Heterogeneous Rock Analysis (HRA) workflow (Fig. 10) provides an important diagnostic check on the stability and validity of the six-cluster electrofacies model applied to the Datta Formation. The curve illustrates the variation in total within-cluster distance across multiple runs of the k-means algorithm. K-means is sensitive to the initial placement of cluster centroids; that's why repeated runs are required in the unsupervised clustering process (MacQueen, 1967; Jain, 2010). A minimal variation in total distance is reflected in a stable model after an optimal number of clusters is selected.

Three important trends are shown in the fall-off plot in this case:

1. A long flat segment between ~15 and ~50 runs indicates that the majority of model initializations converge toward a similar local minimum. This suggests that the six-cluster configuration is statistically stable and not highly sensitive to random initialization.

2. A gradual increase in total distance after ~55 runs reflects occasional centroid placements that lead to less optimal partitioning but remain near the preferred solution. This behavior is common in geological datasets where lithological variability is moderate but not extreme (Doveton, 1994).

3. A sharp rise near the upper-right portion of the plot reflects a small number of poorer clustering solutions. These correspond to rare runs where the initial centroid placement falls into poorly populated portions of the data cloud. Importantly, these higher-distance runs are few, indicating that they do not represent the dominant geological configuration of the dataset.

The behavior of this fall-off plot is consistent with a correctly selected number of clusters for the Datta Formation. In particular:

The stable plateau in the mid-range runs implies that six electrofacies provide a geologically meaningful and statistically repeatable description of the reservoir.

The resulting facies distribution appears consistent with the interbedded sandstone–siltstone–shale succession that characterizes the Datta Formation across the Kohat Plateau (Iqbal et al., 2015).

The small variations seen in the upper-right portion of the plot likely represent thin heterolithic layers and local intercalations, which are well-known in Datta Formation cores and logs. These subtle changes correlate with what is observed in well logs across the plateau.

Because the Datta Formation contains multiple stacked and laterally variable sandstone bodies interbedded with shale, a 6-cluster solution is entirely reasonable. It captures:

- Cemented or compacted intervals (HRA_Facie1).
- Moderately argillaceous sandstones (HRA_Facie_2).
- Clean channel or sheet sandstones (HRA_Facie_3).
- Siltstone units (HRA_Facie_4).
- Transitional heterolithic facies (HRA_Facie_5).
- Shale units (HRA_Facie_6)

This variation matches both the geological expectation and the visual log responses.

5.4 Validation by Confusion Matrix and Accuracy Metrics

To quantitatively assess the performance of the unsupervised electrofacies clustering, a validation against core and cutting-derived facies was conducted using established statistical metrics (Kohavi and Provost, 1998). To evaluate the geological consistency of the classifications computed by Heterogeneous Rock Analysis (HRA), a rigorous framework is provided by the confusion matrix and associated accuracy scores, e.g., Precision, Recall (James et al., 2013). This step is essential to validate the reliability of computed electrofacies and their demonstration as meaningful geological facies before they are applied in models for reservoir characterization.

5.4.1 Data Preparation and Overall Confusion Matrix

The validation dataset comprises core/cutting-derived facies labels (ground truth) and corresponding Heterogeneous Rock Analysis (HRA) electrofacies classifications from 3 wells in the Datta Formation. The HRA Facies are defined as follows for calibration context:

- HRA_Facies 1: Very fine-grained tight sandstone with moderate reservoir quality.
- HRA_Facies 2 & 3: Clean fine- to coarse-grained sandstones with significant reservoir potential.
- HRA_Facies 4: Siltstone/claystone and mixed shale layers.
- HRA_Facies 5 & 6: Heterolithic shaly sands and the primary shale unit of the Datta Formation.

The combined dataset from all three wells contains 4,913 data points for comparison. The overall confusion matrix (Actual vs. Predicted) is presented in Table 2.

5.4.2 Overall Accuracy Metrics

- Overall Accuracy: 79.5% (3906 correct classifications out of 4913).
- Balanced Accuracy: 75.0% (Average of per-class Recall scores).
- Cohen's Kappa (κ): 0.72. This indicates a Substantial level of agreement between the core-based facies and the HRA classification, beyond what would be expected by chance alone.

5.4.3 Precision (by Class)

Precision (Positive Predictive Value) measures the reliability of a specific HRA Facies prediction Table 3. It answers: "When HRA predicts Facies X, how often is it correct?"

Interpretation: HRA Facies 1 (tight sandstone) and 2 (clean sandstone) show excellent precision. Facies 6 (shale) has the lowest precision (~60%), primarily due to misclassifications such as Facies 5 (heterolithic shaly sands), which is geologically reasonable.

5.4.4 Recall (by Class)

Recall (Sensitivity) measures the model's ability to identify all instances of a true facies Table 4. It answers: Of all actual Core

Facies X samples, what proportion did HRA correctly find? Core Facies 1, 4, and 5 are identified with high recall (>81%). The lower recall for Facies 2 (63.5%) is notable; it is most commonly misclassified as HRA Facies 3, suggesting the electrofacies boundary between these two clean sandstone classes may be subtle.

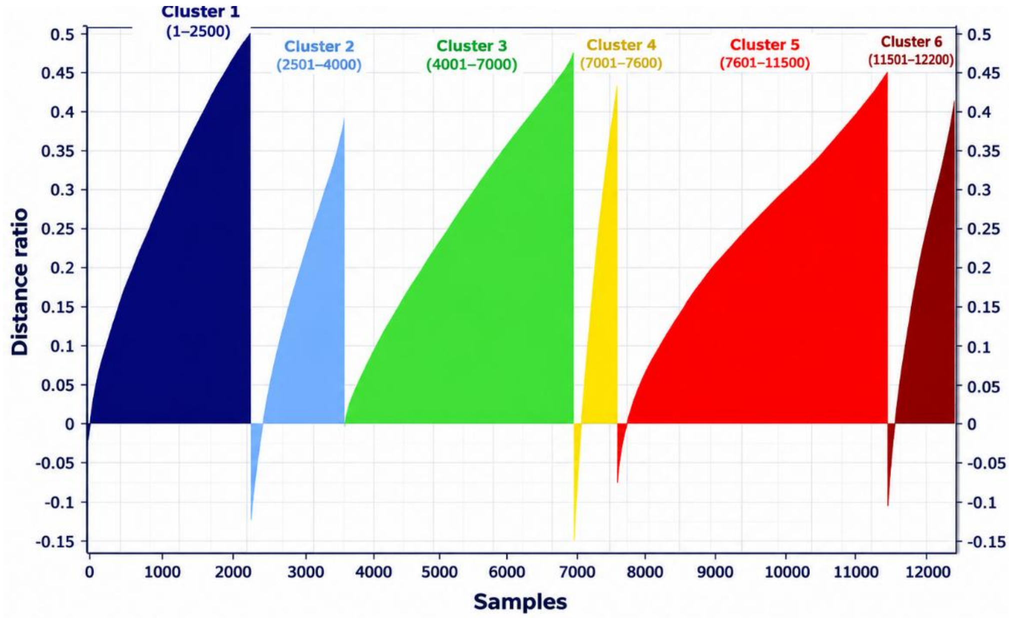

Fig. 9. HRA Silhouette plot of the final choice of 6 clusters. Each color represents a cluster. The distance ratio between the data in a given class and the adjoining class is shown. Negative values are indicative of data that may equally have been assigned to another class.

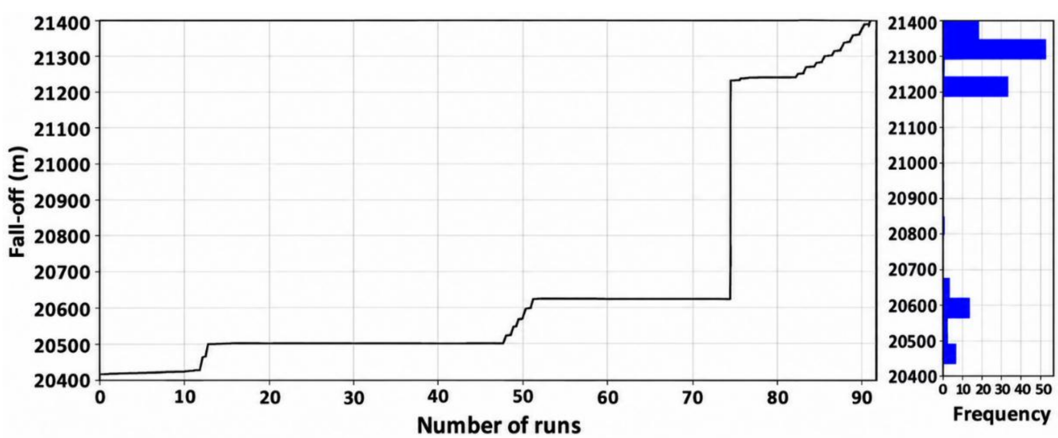

Fig. 10. HRA Fall-off of six clusters. The cluster converged to a distance, which explains that 6 clusters is an adequate number.

Table 2: Overall Confusion Matrix for all Wells.

Actual \ Predicted (HRA)	1	2	3	4	5	6	Σ (Actual)
1 (Core Facies)	**1566**	**10**	**165**	**0**	**20**	**1**	**1762**
2	3	120	53	6	6	1	**189**
3	122	8	708	6	79	5	**928**
4	1	1	5	133	24	0	**164**
5	24	1	46	21	1216	177	**1485**
6	7	0	6	2	97	273	**385**
Σ (Predicted)	**1723**	**140**	**983**	**168**	**1652**	**457**	**4913**

Table 3: Precision by HRA facies class.

HRA Facies	Correct Predictions	Total Predictions	Precision
1	1566	1723	**90.9%**
2	120	140	**85.7%**
3	708	983	**72.0%**
4	133	168	**79.2%**
5	1216	1652	**73.6%**
6	273	457	**59.7%**

Table 4: Recall by core facies class.

Core Facies	Correct Identifications	Total in Class	Recall
1	1566	1762	**88.9%**
2	120	189	**63.5%**
3	708	928	**76.3%**
4	133	164	**81.1%**
5	1216	1485	**81.9%**
6	273	385	**70.9%**

Table 5: Well-specific classification accuracy summary.

Well Name	Data Points	Correct Classifications	Well Accuracy
Makori-1	2748	2109	**76.7%**
Manzalai-2	1000	836	**83.6%**
Maramzai-1	1165	962	**82.6%**
Combined	*4913*	*3907*	***79.5%***

5.5 Well-by-Well Performance Analysis

In well-by-well performance analysis, we are able to quantify the model accuracy separately and then combine the results for the overall accuracy of the model Table 5.

- Makori-1: Shows the lowest but still robust accuracy (76.7%). This well's log interval exhibits the most complex mixing and interbedding of facies, particularly between sandstone (Facies 1/3) and heterolithic/shale units (Facies 5/6), which presents a greater challenge for clustering.
- Manzalai-2: Achieved the highest accuracy (83.6%). The facies distribution in this well appears more distinct, with cleaner separations between reservoir (Facies 2,3,5) and non-reservoir (Facies 4,6) rock types as per the core description.
- Maramzai-1: Performed very well (82.6%). Facies 1 and Facies 5 thick, homogeneous sections are dominant in this well, which are identified with very high consistency by the HRA method and boost the overall score.

5.6 Overall Rating and Analysis of HRA Clustering Results

Very good to excellent performance is demonstrated by HRA electrofacies classification when it is validated alongside the core and cuttings data across the three geographically distinct wells.

- The overall accuracy of 79.5% and a Cohen's Kappa of 0.72 confirm that the clustering workflow has successfully captured the primary geological rock types defined by core description.
- The method is exceptionally reliable in identifying the key reservoir facies (HRA_Facies 1, 2, 3) and the main shale/seal facies (HRA_Facies 5, 6), as evidenced by high precision and recall values for these critical units.
- The observed confusion patterns are geologically interpretable and reasonable. For example, the primary confusion occurs between Facies 5 and 6 (shaly sand vs. shale) and between Facies 2 and 3 (subtle

variations in clean sandstone). These are gradational boundaries in nature, not discrete jumps.

5.7 Weaknesses and Cautions

- Limitations of Cutting Data: The calibration heavily relies on cuttings data, which are depth-averaged and can be contaminated. This likely contributes to the "smearing" effect seen in the confusion between adjacent facies (e.g., 5 vs. 6).
- Class Imbalance: The dataset is dominated by Facies 1 and 5. While the model handles this well, metrics for less frequent facies (like Facies 2 and 4) are based on smaller sample sizes and should be interpreted with caution.
- Inherent Subjectivity in Facies Boundaries: The core facies classification itself has inherent interpreter subjectivity, especially in mixed lithologies. The HRA model provides an objective, reproducible alternative, but the "ground truth" is not absolute.

5.8 Facies Cyclicity and Geological Consistency

The HRA electrofacies scheme shows good agreement with the available core and cuttings data, supporting its application as a reservoir-facies classification tool. The results justify its use for detailed reservoir modeling, stratigraphic analysis, and as a key component in the geological evaluation. The validation results further indicate that the entire facies cycle is repeatedly observed within the logged sections of the Datta Formation. This cyclicity is consistently reproduced by HRA clustering across all three wells, reinforcing the geological credibility of the workflow. The integration of core and predominantly cuttings data within the HRA cluster analysis provides a robust and scalable validation framework for facies classification in data-limited settings.

5.9 Geological Implications and Regional Significance

The selection of a six-electrofacies model aligns with the sedimentological heterogeneity of the Datta Formation, comprising stacked and laterally variable sandstone bodies interbedded with shale and heterolithic units deposited in a fluvial–deltaic to shallow marine setting (Kazmi and Jan, 1997; Shah, 2009; Reading and Collinson, 1996; Miall, 2010). Similar lithofacies variability has been documented in the western Salt Range, reflecting changes in sediment supply and accommodation (Iqbal et al., 2015). In the Kohat Plateau, these depositional processes are further modified by compressional tectonics, producing pronounced lateral and vertical heterogeneity.

HRA analysis confirms this framework, with HRA_Facies 2 and 3 exhibiting the highest reservoir potential in terms of thickness, cleanliness, and porosity-permeability trends, while HRA_Facies 1 shows limited potential and HRA_Facies 4 functions as a lateral and vertical flow barrier. The repeated layers of heterolithic shaly sands are categorized as HRA_Facies 5, and HRA_Facies 6 represents the shale unit in the Datta Formation in the Kohat area. The results are presented in the integrated plot (Fig. A3) of the Manzalai-2 well.

Track 1 shows the depth interval.2: Gamma Ray; Track 3: volume of shale. 4-5 show resistivity, Neutron Density & Sonic data. Tracks.6,7,8,9 shows Saturation of water, total & effective porosities, movable hydrocarbon, & log-based permeability. Track .10 shows lithology volumes, net, reservoir & rock flags. Track 11 shows the cumulative heterogeneity volumetric curves. Tracks. 12-14 show the Core-Cuttings Facies flag & Facies code. Tracks 15-16 show the Electrofacies from K-means clustering and assigned codes. The correlation of all seven wells is presented in Figure A4, where the facies trend can be seen on a regional scale.

The observed vertical stacking and lateral continuity underscore the structural and depositional controls on reservoir distribution. Together, the electrofacies framework and HRA results provide a robust basis for correlating litho-units, identifying net-pay zones, and mapping heterogeneity, thereby supporting targeted exploration and production strategies within the Kohat fold-and-thrust belt.

6 Conclusion

This study presents the first basin-scale integration of conventional petrophysical analysis and machine-learning–based electrofacies classification for the Jurassic Datta Formation across the Kohat Plateau, Northwest Pakistan, using wireline log data from seven wells. Conventional petrophysical evaluation successfully delineates the primary lithological units and quantifies key reservoir parameters, revealing strong spatial variability in reservoir quality across the basin, with productive, coarser-grained sandstone facies dominating the southern wells and tighter, fine-grained facies prevalent in the northern and central sectors.

The PCA-assisted k-means clustering implemented through Heterogeneous Rock Analysis (HRA) identifies six electrofacies that capture the internal heterogeneity of the Datta Formation, including clean sandstones, argillaceous sands, siltstones, shales, and compacted intervals. These electrofacies show strong consistency with log signatures, core and cuttings descriptions, and observed reservoir behavior, confirming their geological relevance. Model performance was evaluated using a confusion matrix and accuracy metrics, yielding an overall classification accuracy of 79.5%, which demonstrates the robustness and reliability of the electrofacies framework.

Overall, the integrated petrophysical–electrofacies workflow improves reservoir characterization by reducing interpretational uncertainty, enhancing lateral correlation, and providing a quantitative basis for reservoir zonation and development planning. The methodology is directly applicable to exploration and development workflows in the Kohat Basin and other structurally complex fold-and-thrust belts.

Authors' Contributions

Syed Asad Ali Kazmi, conceived and designed the study, conducted petrophysical and electrofacies analyses, interpreted the results, and prepared the original manuscript draft. He also performed data integration, statistical evaluation, and visualization of the results. Nosheen Akhtar, contributed to the research framework, methodological review, and critical revision of the manuscript. She provided technical guidance during the interpretation stage, reviewed the final manuscript for intellectual content, and approved the submitted version for publication.

Acknowledgement

For this study, wireline logs, core, and cuttings data were provided by the Directorate General of Petroleum Concession of Pakistan (DGPC).

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

Ahmad, S., Ali, F., Sayab, M., Ahmad, I., & Hamidullah, S. (1999). Structural geometry of the Himalayan Frontal Thrust Zone: Surghar Range, Pakistan. *Journal of Himalayan Earth Sciences*, *32*(1), 13–23. http://ojs.uop.edu.pk/jhes/article/view/1517

Asquith, G., & Krygowski, D. (2004). *Basic well log analysis* (2nd ed.). American Association of Petroleum Geologists.

Beck, R. A., Burbank, D. W., Sercombe, W. J., Khan, A. M., & Lawrence, R. D. (1996). Late Cretaceous ophiolite obduction and Paleocene India–Asia collision in the westernmost Himalaya. *Geodinamica Acta*, *9*(2–3), 114–144. https://doi.org/10.1080/09853111.1996.11105281

Bilal, A., Yang, R., Janjuhah, H. T., Mughal, M. S., Li, Y., Kontakiotis, G., & Lenhardt, N. (2023). Microfacies analysis of the Palaeocene Lockhart limestone on the eastern margin of the Upper Indus Basin (Pakistan): Implications for the depositional environment and reservoir characteristics. *The Depositional Record, 9*(1), 152–173. https://doi.org/10.1002/dep2.222

Catuneanu, O. (2019). *Principles of sequence stratigraphy*. Elsevier.

Cherana, A., Aliouane, L., Doghmane, M. Z., Ouadfeul, S.-A., & Nabawy, B. S. (2022). Lithofacies discrimination of the Ordovician unconventional gas-bearing tight sandstone reservoirs using a subtractive fuzzy clustering algorithm applied to the well log data: Illizi Basin, the Algerian Sahara. *Journal of African Earth Sciences, 196*, 104732. https://doi.org/10.1016/j.jafrearsci.2022.104732

Dani, T. M., Khan, M. A., & Kakar, M. A. (1995). Structural style and stratigraphy of the Kohat Basin. *Geological Bulletin, University of Peshawar, 28*, 39–50.

Davis, J. C. (2018). Electrofacies in reservoir characterization. In B. D. S. Sagar, Q. Cheng, & F. P. Agterberg (Eds.), *Handbook of mathematical geosciences*. Springer. https://doi.org/10.1007/978-3-319-78999-6_11

Doghmane, M. Z., Ouadfeul, S.-A., Benaissa, Z., & Eladj, S. (2022). Classification of Ordovician tight reservoir facies in Algeria using neuro-fuzzy algorithm. In S.-A. Ouadfeul (Ed.), *Artificial intelligence and heuristics for smart energy efficiency in smart cities: Case study: Tipasa, Algeria* (pp. 889–895). Springer. https://doi.org/10.1007/978-3-030-92038-8_91

Doveton, J. H. (1994). *Geologic log analysis using computer methods* (Vol. 2). Published by the American Association of Petroleum Geologists.

Ellis, D. V., & Singer, J. M. (2007). *Well logging for earth scientists* (2nd ed.). Springer.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Fatmi, A. N. (1973). *Lithostratigraphic units of the Kohat–Potwar Province, Indus Basin, Pakistan* (Geological Survey of Pakistan Memoir No. 10). Geological Survey of Pakistan.

Ghani, H., Ahmad, I., Ahmad, S., Ali, F., & Irum. (2015). Structural evolution of southern Kohat fold and thrust belt, a case study from Karak area, Pakistan. *Journal of Himalayan Earth Sciences, 48*(2), 1–8. http://ojs.uop.edu.pk/jhes

Guo, Q.-J., Strauss, H., Liu, C.-Q., Zhao, Y.-L., Pi, D.-H., Fu, P.-Q., Zhu, L.-J., & Yang, R.-D. (2005). Carbon and oxygen isotopic composition of Lower to Middle Cambrian sediments at Taijiang, Guizhou Province, China. *Geological Magazine, 142*(6), 723–733. https://doi.org/10.1017/S0016756805001202

Hartigan, J. A., & Wong, M. A. (1979). A K-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 28*(1), 100–108. https://doi.org/10.2307/2346830

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*(6), 417–441 https://doi.org/10.1037/h0071325

Iqbal, S., Jan, I. U., Akhter, M. G., & Bibi, M. (2015). Palaeoenvironmental and sequence stratigraphic analyses of the Jurassic Datta Formation, Salt Range, Pakistan. *Journal of Earth System Science, 124*(4), 747–766. https://doi.org/10.1007/s12040-015-0572-y

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys, 31*(3), 264–323. https://doi.org/10.1145/331499.331504

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer. https://doi.org/10.1007/b98835

Kadri, I. B. (1995). *Petroleum geology of Pakistan*. Pakistan Petroleum Limited.

Kafayatullah, K., Arif, M., & Shah, M. T. (2006). Petrography of sandstones from the Kamlial and Chinji formations, southwestern Kohat Plateau, NW Pakistan: Implications for source lithology and paleoclimate. *Journal of Himalayan Earth Sciences, 39*, 1–13. http://ojs.uop.edu.pk/jhes

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. Wiley. https://doi.org/10.1002/9780470316801

Kazmi, A. H., & Jan, M. Q. (1997). *Geology and tectonics of Pakistan*. Graphic Publishers.

Khan, N., & Rehman, K. (2019). Petrophysical evaluation and fluid substitution modeling for reservoir depiction of Jurassic Datta Formation in the Chanda oil field, Khyber Pakhtunkhwa, northwest Pakistan. *Journal of Petroleum Exploration and Production Technology, 9*(1), 159–176. https://doi.org/10.1007/s13202-018-0513-9

Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning, 30*(2–3), 271–274. https://doi.org/10.1023/A:1017181826899

Kumar, B., & Kishore, M. (2006). *Electrofacies classification—A critical approach.* In *6th International Conference and Exposition on Petroleum Geophysics*, Kolkata, India (pp. 822–825).

Lee, S. H., Khargoria, A., & Datta-Gupta, A. (2002). Electrofacies characterization and permeability predictions in complex reservoirs. *SPE Reservoir Evaluation & Engineering, 5*(3), 237–248. https://doi.org/10.2118/78662-PA

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California, Berkeley.

Mahmood, M. N., & Guo, B. (2019). An analytical method for optimizing fracture spacing in shale oil reservoirs. *SPE Liquids-Rich Basins Conference–North America*. Society of Petroleum Engineers. https://doi.org/10.2118/197083-MS

Meissner, C. R., Master, J. M., Rashid, M. A., & Hussain, M. (1974). *Stratigraphy of the Kohat Quadrangle, Pakistan* (U.S. Geological Survey Professional Paper 716-D). U.S. Geological Survey. https://doi.org/10.3133/pp716D

Miall, A. D. (2010). *The geology of fluvial deposits: Sedimentary facies, basin analysis, and petroleum geology*. Springer.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2*(11), 559–572. https://doi.org/10.1080/14786440109462720

Pivnik, D. A., & Sercombe, W. J. (1993). Compression- and transpression-related deformation in the Kohat Plateau, NW Pakistan. In P. J. Treloar & M. P. Searle (Eds.), *Himalayan tectonics* (Geological Society, London, Special Publications, Vol. 74, pp. 559–580). Geological Society, London. https://doi.org/10.1144/GSL.SP.1993.074.01.37

Posamentier, H. W., & Walker, R. G. (Eds.). (2006). *Facies models revisited* (SEPM Special Publication No. 84). Society for Sedimentary Geology. https://doi.org/0.2110/pec.06.84

Radhy, J., & Nasser, M. (2017). Role of cluster analysis in logfacies and depositional environments recognition from well log response for Mishrif Formation in southeast Iraq. *International Journal of Engineering and Advanced Research, 3*. https://doi.org/10.25125/engineering-journal-IJOER-DEC-2017-5

Reading, H. G., & Collinson, J. D. (1996). *Clastic sedimentary environments*. Blackwell Science.

Rider, M. H. (1996). *The geological interpretation of well logs* (2nd ed.). Gulf Publishing.

Rojas, L. J., Tveritnev, A., & Pinillos, C. E. (2020). A multidisciplinary approach for rock typing characterization in a highly heterogeneous carbonate reservoir in Abu Dhabi, UAE [Poster presentation]. *2020 AAPG Middle East Region Geoscience Technology Workshop, 3rd Edition: Carbonate Reservoirs of the Middle East*, Abu Dhabi, United Arab Emirates.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Schlumberger. (1989). *Log interpretation principles/applications*. Schlumberger Wireline & Testing.

Serra, O. (1984–1986). *Fundamentals of well-log interpretation* (P. Westaway & H. Abbott, Trans.; Vols. 1–2). Elsevier.

Serra, O., & Abbott, H. T. (1982). The contribution of logging data to sedimentology and stratigraphy. *SPE Journal*, *22*(1), 117–131. https://doi.org/10.2118/9270-PA

Shah, S. M. I. (2009). *Stratigraphy of Pakistan* (Geological Survey of Pakistan Memoirs No. 22). Geological Survey of Pakistan.

Tiab, D., & Donaldson, E. C. (2015). *Petrophysics: Theory and practice of measuring reservoir rock and fluid transport properties* (3rd ed.). Gulf Professional Publishing.

Willis, B. (2013). *Heterogeneous Rock Analysis (HRA) technical notes*. Schlumberger.

Wandrey, C. J., Law, B. E., & Shah, H. A. (2004). *Patala-Nammal composite total petroleum system, Kohat-Potwar geologic province, Pakistan* (U.S. Geological Survey Bulletin 2208-B). U.S. Geological Survey. https://doi.org/10.3133/b2208B

Yeats, R. S., & Thakur, V. C. (2008). Active faulting south of the Himalayan front: Establishing a new plate boundary. *Tectonophysics*, *453*(1–4), 63–73. https://doi.org/10.1016/j.tecto.2007.06.017

Appendix-A

Fig. A1. Example of a computer-processed interpretation (CPI) plot showing multi-mineral modeling results for the Chanda Deep-1 well. Track 1 shows depth; Track 2, gamma ray (GR); Track 3, shale volume (Vsh); Tracks 4–5, resistivity, neutron-density, and sonic logs; Tracks 6–9, water saturation (Sw), total and effective porosities, movable hydrocarbon, and permeability; and Tracks 10–13, lithology volumes, net, reservoir, and rock flags.

Fig. A2. Correlation of study wells across the Kohat Plateau, from Mramzai-1 in the north to Chanda Deep-1 in the south, showing the distribution and correlation of the Datta Formation.

Fig. A3. Example of a computer-processed interpretation (CPI) plot showing multi-mineral modeling results for the Manzalai-2 well. Quartz with clay dominates the upper part of the Datta Formation, whereas cleaner sand is interpreted in the middle and lower parts. The panel also shows cumulative heterogeneity volumetric curves, core- and cutting-derived facies flags and codes, and electrofacies classifications generated using k-means clustering with corresponding facies codes.

Fig. A4. Correlation panel integrating petrophysical and electrofacies analyses of the study wells in the Kohat Basin.