International Journal of Multiphysics
Volume 18, No. 4, 2024
ISSN: 1750-9548

Comparing Facies Prediction Performance of Machine Learning Models Trained on Well and Core Data: A Case Study from Lower Indus Basin

1Noor ul huda choudhry, 2Muhammad Ali Tahir, 3Asad Taimur,

1PhD , Melbourne college of earth and energy, the University of Oklahoma.

Email: Noor.u.choudhry-1@ou.edu

2Faculty member, Institute of Geographical Information Systems, National University of Science and Technology.

Email: ali.tahir@igis.nust.edu.pk

3Assistant Manager (Geophysics), Mineral Exploration & Development Organization, Pakistan

Email: asad07.t@live.com

Abstract

Machine learning has excellent potential to predict rock types and depositional trends at a sub-centimeter scale using borehole data in oil and gas wells. The required dataset includes core plug extracted from the wells and well-log data acquired through different tools in run in boreholes from six wells in the lower Indus basin. Core plugs are the only subsurface data that is true to geologic scale and inherent heterogeneity. The research employs a rock-type driven labeling scheme and a rock depositional process focused classification scheme to interpret the training data from core plugs at a sub-centimeter scale. To generate predictions for lithology and facies, an “RGB log” (RGBL) is developed to summarize the core plug image at each depth step. The use of RGBL data has generated even more accurate results and requires far less computing power than core image data. On the other hand, it is anticipated that well-log data will continue to be inadequate in predicting rock types or depositional trends at the sub-centimeter level due to logging speed and step interval. To overcome this challenge, multiple curves are used as inputs for activation functions to predict rock types from well-log data with signatures of encountered rock types. The study demonstrates the potential to transform large quantities of photographed core into a normalized digital format for geologic insights. The methodology involves a machine learning workflow developed in python; employed for the analysis of core image data in a scalable and reproducible manner. This approach can be extended to other geologic basins with similar clastic depositional trends, provided there is an abundance of photographed core plugs. RandomForest and GradientBoosting were used to estimate the facies using well log data; RandomForest was slightly higher in accuracy at 87.1% compared to GradientBoosting's 85.7%. Using RGB log data, MLP-SVM predicted facies with an overall accuracy of 92.31%, with metrics for precision, recall, and F1 scores of 0.96, 0.93, and 0.94, respectively.

Introduction

The lower Indus basin still has huge upside gas potential. With exploration efforts dating back to 1980’s, Miano Gas Field was discovered at B-Sand Interval of the early Cretaceous Lower Goru Formation in August 1994 and is something of a geological puzzle. The sands which were supposed to flow gas proved to be either low porosity (tight) or produced water. This relates directly to facies variation within the Lower Goru Formation. With only B-

Sand interval producing commercially viable gas, other sand bodies were extensively tested but due to a poor understanding of depositional geomorphology, most models were inconclusive in evaluating the true potential of Lower Goru Formation.

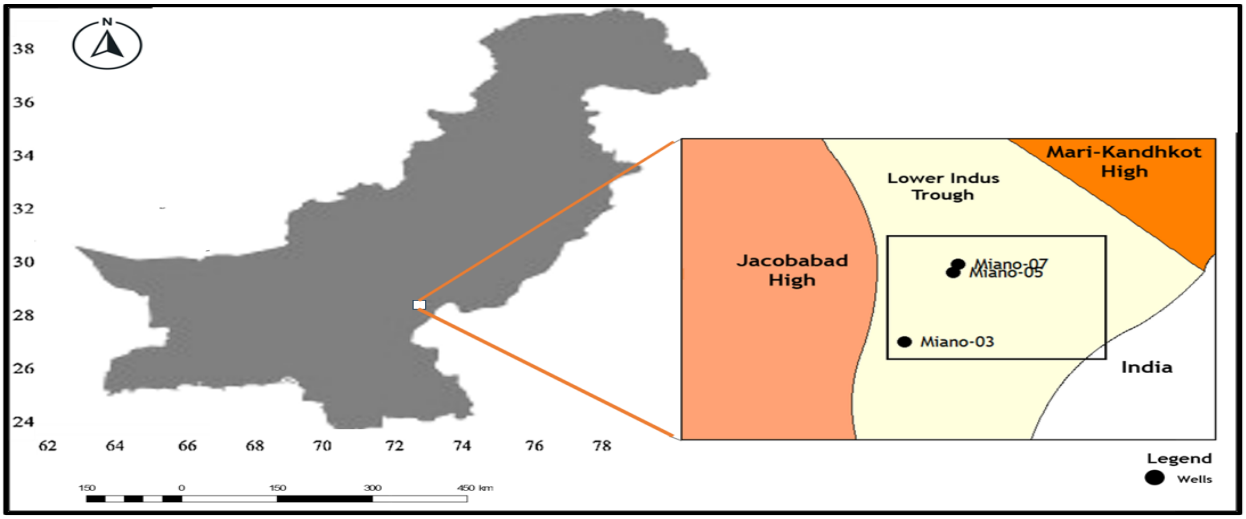


Figure 1 Study area map

However, the B-Sand interval has been exploited thoroughly for commercial gas flow for the past three decades. A lack of facies control and understanding in the deltaic and pro-deltaic environment have hindered subsequent exploration endeavors. This is precisely where machine learning driven workflows with a high level of detail can be of immense value.

Identifying facies (the characteristics of a rock that reflect its origin and differentiate it from other rocks in its vicinity) within the morphologies is the key to understand how the hydrocarbon reservoir will perform throughout its life. The main challenges include lack of control over facies identification, interpretation bias, uncertainties in data acquisition and the level of detail is far less than required for a thorogh distinction depositional patterns. With the advent of computing power and machine learning workflows, these challenges can be overcome to a significant degree in understanding the depositional patterns, heterogeneity and the associated uncertainty. The unprecedented capability of machine learning models to provide unbiased identification and classification with scaleable accuracy is leveraged in this study to assess the performance of a data driven machine learning in its ability to create value in facies identification at a much finer scale than what the current technology can offer.

Machine learning workflows have been used for hydrocarbon reservoir evaluation in the oil and gas sector for over forty years. Recently, there has been a resurgence of machine-learning focused research for subsurface characterization. The historical progression of integrating machine learning applications into geoscience workflows represents a gradual shift from conventional manual approaches to more advanced computational methods, fundamentally altering our understanding and subsurface interpretation. In its early stages, geoscientists heavily relied on traditional methods for data analysis and interpretation, grappling with the complexities of large and intricate datasets inherent to the field. The introduction of machine learning into geoscience gained momentum as researchers sought computational solutions to address these challenges. Initial applications centered on basic pattern recognition and statistical analyses, laying the groundwork for the development of more sophisticated algorithms. The adoption of supervised learning methods, as such decision trees and support vector machines marked a notable advancement, automating tasks such as facies prediction from well logs. The subsequent emergence of deep learning, specifically with convolutional neural networks (CNNs) and recurrent neural networks (RNNs), further transformed geoscience workflows, allowing models to capture intricate spatial and temporal relationships within geological data. Over time, interdisciplinary collaboration has gained prominence, combining domain expertise with machine learning capabilities to enhance the accuracy and reliability of predictions in geoscience applications.

The Society of Exploration Geophysicists organized a machine learning hackathon in 2016 poised at predicting rock types and depositional trends from a core-calibrated well-log dataset. The most accurate model utilized a boosted tree approach with a median accuracy of 0.64 from nine classes. Encouraged by the promising results, this study uses an "RGB log" (RGBL) and core images at each data acquired step interval. This method needs a lot less computational power than core image data and predicts results with a higher accuracy. The WaveNet model was proposed to predict individual lithology classes of reservoir potential clastic facies.

Although wireline logs are useful to oil and gas professionals, they are limited in their resolution and cannot fully capture the subsurface depositional character, especially at fine scales (<5 cm). This is where core data becomes important. Machine learning algorithms are now being used to predict lithology and facies in cored well images at the centimeter scale.

Presently, machine learning has become an important component of geoscience, contributing to more efficient resource exploration, environmental monitoring, and informed decision-making in the face of complex geological challenges.

Literature Review

Clastic depositional environments are diverse, heterogeneous and have massive potential for being hydrocarbon reservoirs. The study focuses on the potential of automated lithology prediction in clastic depositional conditions using borehole data from the Miano block in the Lower Indus Basin. Facies prediction is a critical yet challenging endeavor for oil and gas field development, particularly at centimeter scale. Since the deposition does not happen randomly and depends entirely on the energy of water at different stages of sea level rise and fall, there are patterns in nature (Ahmad et al., 2012). The effort to integrate core-based facies to reservoir scale models is especially complicated while trying to capture the thin-bedded heterogeneity that is common to deposition via interacting dynamic processes (wave energy, tidal regime, currents etc.; each with different energy), which re-work and disperse fluvial clastic sediments in a deltaic setting (Wolf et al., 1982b). Depositional morphology in deltaic environments include distributary channels, river-mouth bars, inter-distributary bays, tidal flats, shore-face, beaches, swamps, marshes, and evaporite flats (Fu et al., 2012).

Identifying facies (the characteristics of a rock that reflect its origin and differentiate it from other rocks in its vicinity) within the morphologies is the key to understanding how the hydrocarbon reservoir will perform throughout its life (Baldwin et al., 1990). The main challenges include lack of control over facies identification, interpretation bias, uncertainties in data acquisition and the level of detail are far less than required for a thorough distinction depositional patterns. With the advent of computing power and machine learning workflows, these challenges can be overcome to a significant degree in understanding the depositional patterns, heterogeneity, and the associated uncertainty (Hall, 2016). The unprecedented capability of machine learning models to provide unbiased identification and classification with scalable accuracy is leveraged in this study to assess the performance of a data driven machine learning in its ability to create value in facies identification at a much finer scale than what the current technology can offer.

The subsequent emergence of deep learning, specifically with convolutional neural networks (CNNs) and recurrent neural networks (RNNs), further transformed geoscience workflows, allowing models to capture intricate spatial and temporal relationships within geological data (Quinlan, 1986). Over time, interdisciplinary collaboration has gained prominence, combining domain expertise with machine learning capabilities to enhance the accuracy and reliability of predictions in geoscience applications (Liu et al., 2018).

Advances in machine learning have enabled the use of semi-supervised learning and self-training methods in geoscience applications. Dunham et al. (2020) showed how these techniques can improve well-log classification, increasing the accuracy and efficiency of lithology prediction. Additionally, integrating subsurface core images with depth-registered datasets has opened new paths for lithology prediction and characterization. Meyer et al. (2020) introduced CoreBreakout, a novel approach to generate depth-registered datasets from subsurface core images, providing valuable insights into lithological properties and depositional patterns.

The use of machine learning in geoscience has also helped predict lithology and rock layers in cored wells at a very detailed, centimeter-level scale. A study by Martin et al. (2021) showed that machine learning techniques can accurately predict these fine-scale features, allowing for better understanding of underground reservoir complexities. These advancements demonstrate the increasing significance of machine learning in geoscience and its potential to transform subsurface analysis and forecasting.

In the world of digital rock analysis, Luthi (1994) presented a method to divide digital rock images into distinct bedding layers. This approach used texture energy and cluster labels to segment the images. This has opened the door for more advanced and accurate characterization of reservoir rocks. This allows for better understanding and prediction of the rock's properties and how it was formed. By combining these innovative techniques with machine learning, the industry can now do more precise and detailed analyses of what's underground. This leads to improved reservoir characterization and better oil and gas exploration.

In the digital rock analysis field, Luthi (1994) developed a textural segmentation method. This technique uses texture energy and cluster labels to divide digital rock images into bedding units. This approach has paved the way for more advanced and precise characterization of reservoir rocks. It enables better understanding and prediction of lithological properties and depositional environments. By combining these innovative techniques with machine learning algorithms, the industry can achieve more accurate and detailed subsurface analyses. This ultimately leads to improved reservoir characterization and hydrocarbon exploration.

The Lower Indus Basin is surrounded by the Central Indus and Sulaiman Fold-Belt Basins to the north and the Kirthar Fold-Belt Basin to the west. This area contains both clastic and carbonate sediments dating from ancient times to the present (Fig. 2). Located within the "Indus Platform and Foredeep" tectonic region, it has multiple structural zones characterized by tilted fault blocks and thrust-fault anticlines (Kazmi and Jan 1997; Nazir and Fazeelat 2014). During the early Cretaceous period, the Indian Plate shifted northward into warmer areas (Jadoon et al. 1992; Kazmi and Jan 1997; Khalid et al. 2014a; Fig. 3a). Along the western shelf, the Lower Cretaceous Sembar and Goru formations, consisting of marine shales, limestone, and nearshore sandstones, were deposited over the Sulaiman Limestone Group, under a widespread erosional layer. This area transitioned into sedimentary rocks called sandstones, including the Lumshiwal and Pab formations in the west and the Tura Formation in the east (Khalid et al. 2014b). As the Indian Plate moved northward towards the Asian Plate in the Late Cretaceous, the formation of the Bengal Basin seafloor led to the buildup of flysch around the Indian Plate (Shah 2009; Fig. 3b). Later, a significant flooding event occurred, followed by the widespread deposition of the Upper Goru Formation, which acted as a regional seal for the reservoirs of the Lower Goru Formation (Sahito et al. 2013).

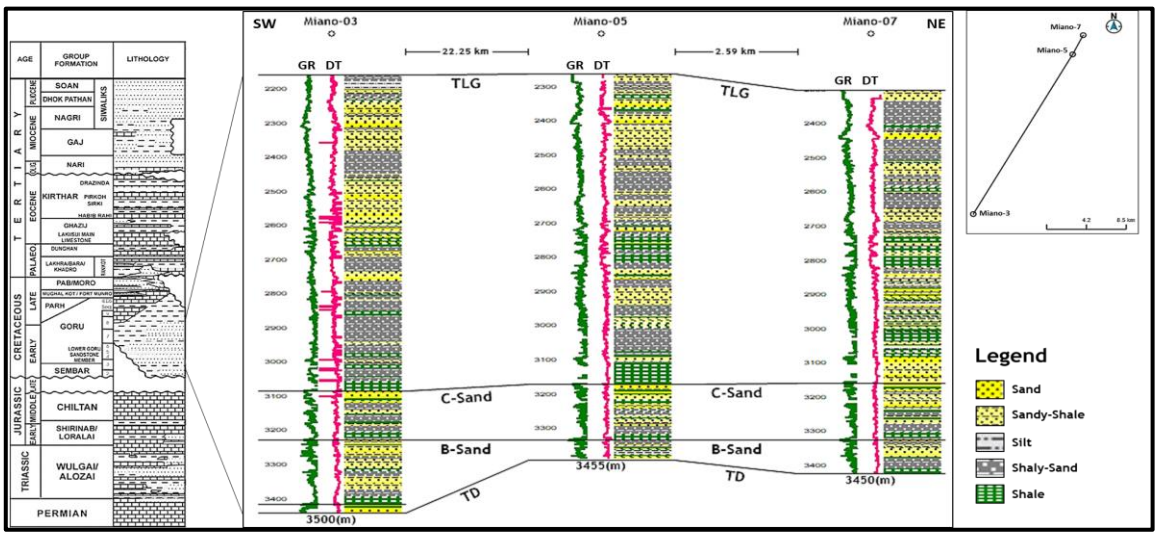
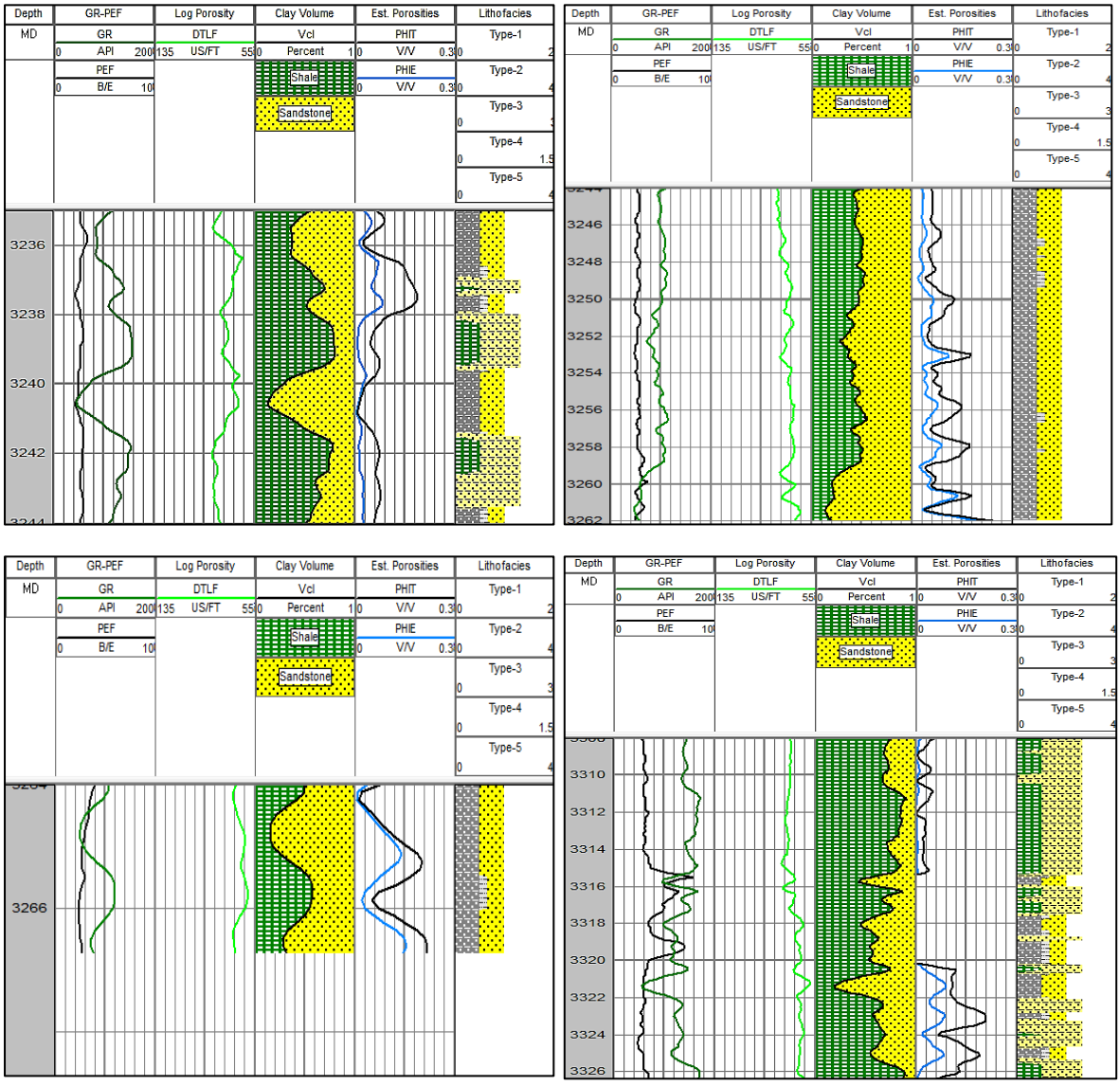


Figure.2 Sequence stratigraphic correlation of Miano-03, Miano-05, Miano-07

Methodology

International Journal of Multiphysics
Volume 18, No. 4, 2024
ISSN: 1750-9548

Lithofacies are predicted using well logs and RGB pixel information of core images as different lithology has different colors for example sandstone common color are tan, brown, yellow, red, grey, pink, white based on composition, sandstone can be classified based on RGB response (Passay et al., 2006). Open-hole log interpretation of the wells Miano 3, 5 and 7 classified lithology of the reservoir into 5 lithotypes Lithotype 1 was Sand, Lithotype 2 was Sandy Shale, Lithotype 3 was Silt, Lithotype 4 was Shaly-sand and Lithotype 5 was Shale. Well data was preprocessed NaN values were removed in training and testing data and the data was merged into a single data-frame. Skewness was removed in curve data required through machine learning algorithms and the dataset was normalized using yeo-johnson normalization. The outliers were subsequently removed. Miano 5 was used as a blind well and two machine learning algorithms (Random Forest Classification and Gradient Boosting Classification) were used to train wells Miano-3 and Miano-7 and predict facies on Miano-5 (Chen et al., 2016). Random Forest predicted facies with little more accuracy than Gradient Boosting. Randon Forest classifier had an accuracy of 87.1% and accuracy of each lithology column were Sand 92%, Sandy Shale 88%, Siltstone 87.16%, Shaly Sand 88% and Shale 88%. Gradient Boosting had an accuracy of 85.7% and accuracy of each lithology column were Sand 91%, Sandy Shale 88%, Siltstone 87.14%, Shaly Sand 86% and Shale 88%.



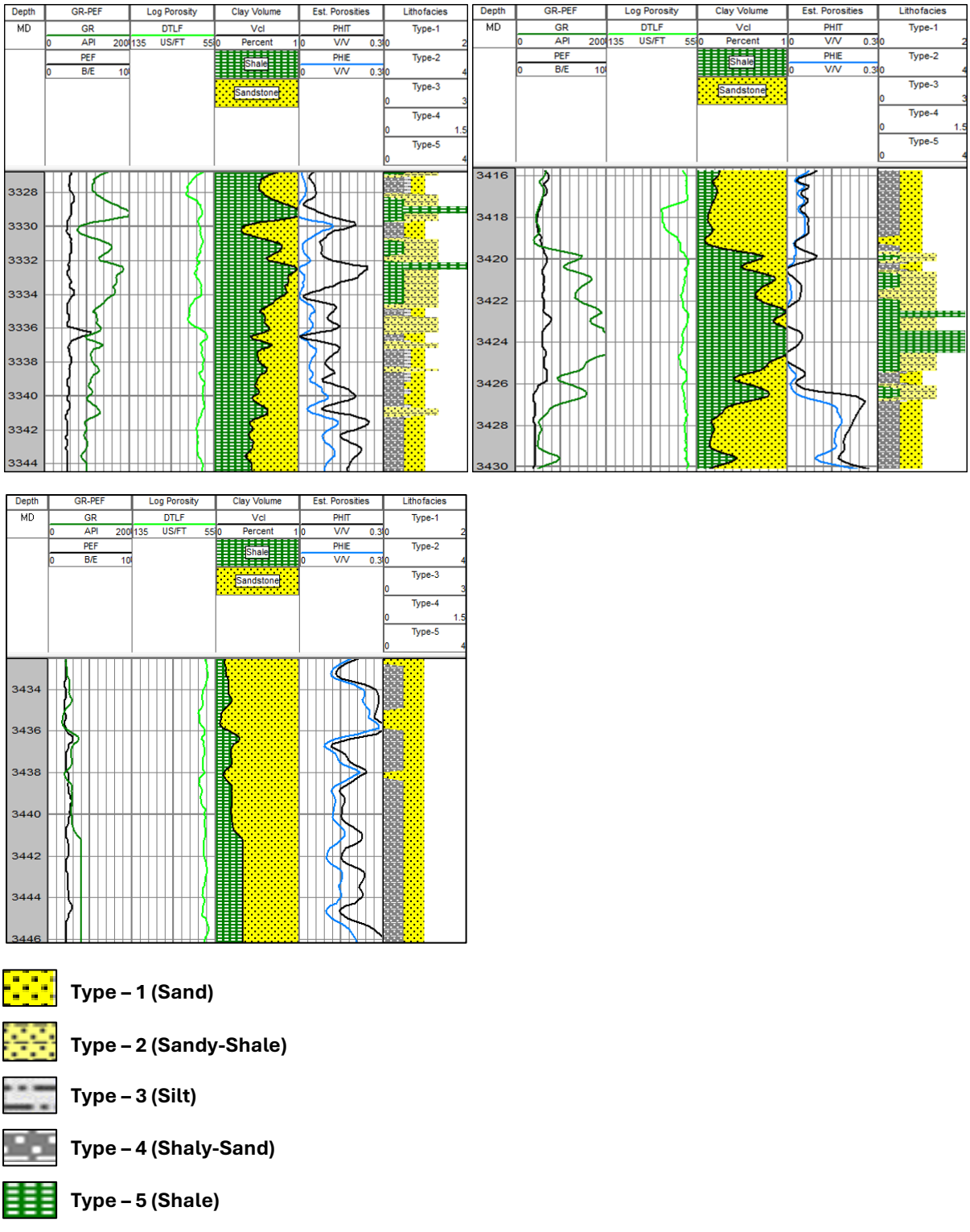
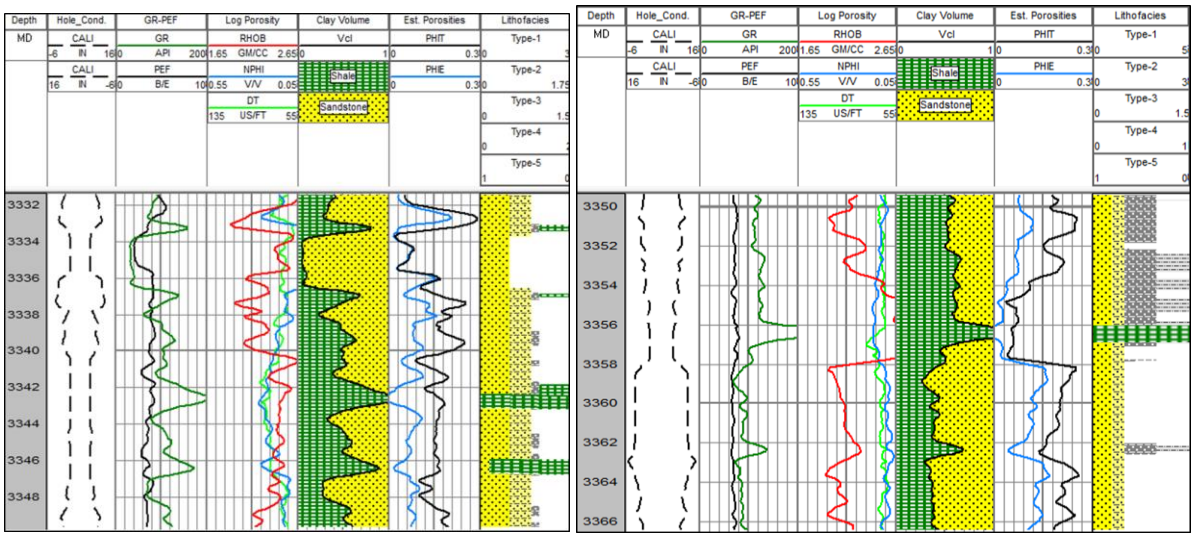


Figure 3 Manual interpretation of litho facies of Miano-3



Type – 1 (Sand)

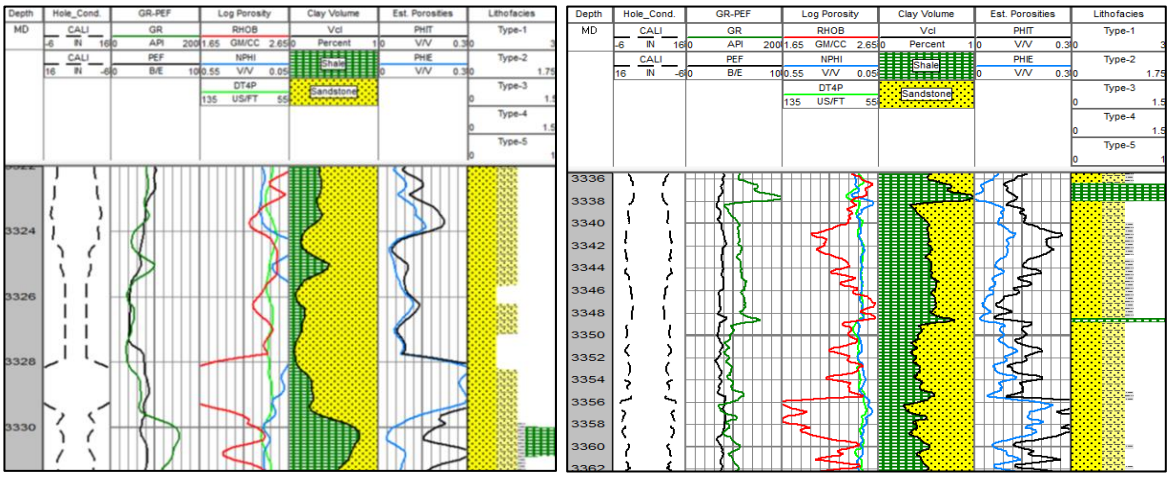
Type – 2 (Sandy-Shale)

Type – 3 (Silt)

Type – 4 (Shaly-Sand)

Type – 5 (Shale)

Figure 4 Manual interpretation of litho facies of Miano-5



Type – 1 (Sand)

Type – 2 (Sandy-Shale)

Type – 3 (Silt)

Type – 4 (Shaly-Sand)

Type – 5 (Shale)

Figure 5 Manual interpretation of litho facies of Miano-7

Core images were split into RGB components the Red, Green and Blue pixel rows were average to plot it as a pseudo litho-log, yeo-johson normalization was used to normalize the RGB logs from 0-255 to 0-1 per channel (red green blue) representing the mean value for each channel across the core (Hall and Hall, 2017). the RGB logs capture fine-scale detail that the GR log does not. MLP-SVM (Multilayer Perception Support Vector Machines) was used to predict facies using RGB log instead the GR log. Miano-5 was a blind well and Miano 3, Miano-7 are used as training dataset.

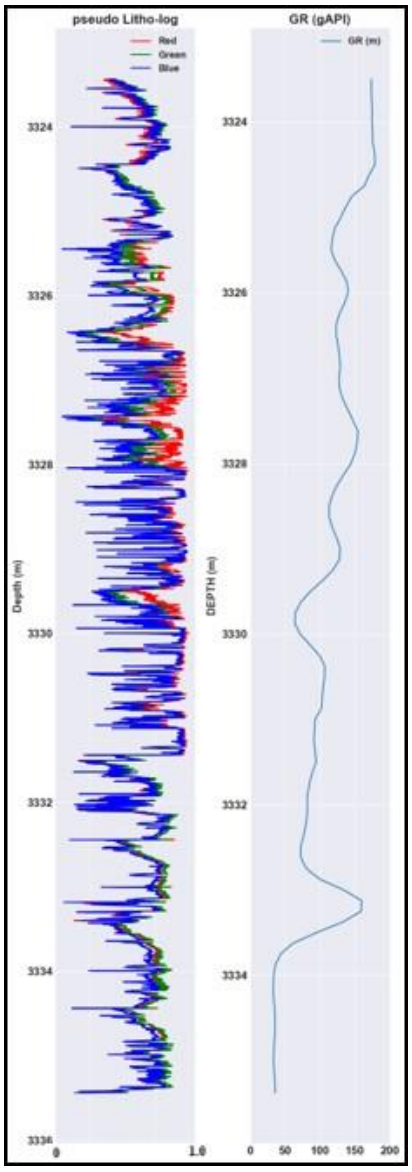


Figure 6 An illustration of the well Mian-5, RGB log (left), and gamma ray log (GR, right) (depth in meters). To represent the mean value for each channel across the core, we normalize the RGB log's scale from 0–255 values to 0–1 on a per-channel basis (i.e., red, green, blue, and gray). Keep in mind that the GR log lacks fine-scale detail, but the RBG log does.

According to confusion metrix Class 2: 13 instances were correctly classified as class 2, and 1 instance of class 6 was misclassified as class 2 Class 4: 12 instances were correctly classified as class 4, and 2 instances of class 5 were misclassified as class 4. The overall accuracy of 92.31% and performance metric showed precision of 0.96, recall of 0.93 and F1 score of 0.94.

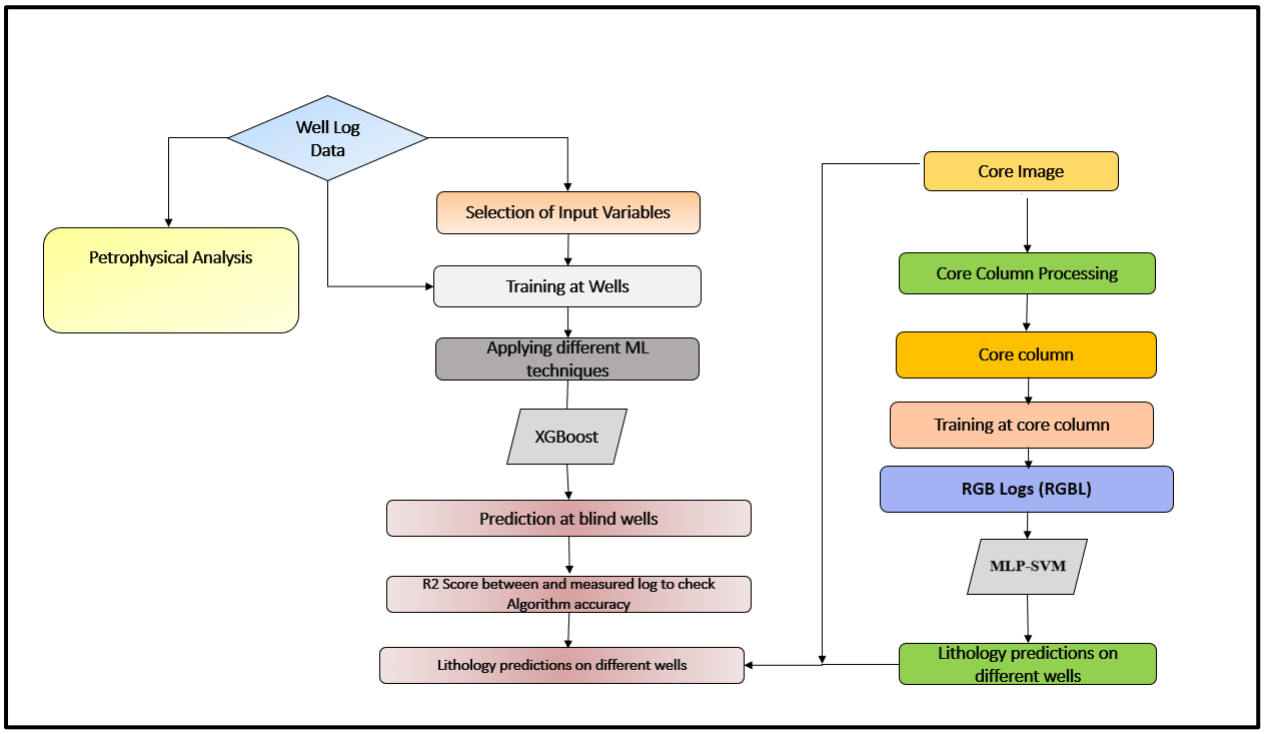


Figure 7 Workflow of the study

Discussion and Results

Core sedimentology was evaluated and seven lithotypes have been defined, mainly based on the different grain size and amount of sedimentary and biogenic structures. Lithotype 1A is tight, very coarse-grained, siderite-cemented, strongly bioturbated, shaly sandstone rich in mollusk fragments. Lithotype 1B is very coarse-grained, strongly bioturbated, shaly sandstone showing common biogenically distributed granules to pebbles and occasional mollusk fragments. Determinable burrow structures comprise mainly Ophiomorpha sp. traces. Lithotype 2A is characterize as Whitish, coarse-grained, quartz-cemented sandstone to granule-sized conglomerate with occasional. scattered pebbles. Recognizable bedding types are represented by low- to high-angle cross-bedding. Fore-set beds are partly accentuated by thin layers of granules to pebbles. Remarkable is the frequent occurrence of rip-up clay clasts. This lithotype displays further occasional Ophiomorpha sp. burrow structures and contains a thin interlayer of fine-medium-grained sandstone showing a single large coalified wood fragment (>10 cm). Also visible are isolated pyrite nodules. Lithotype 2B is recognized as Whitish to light grey, fine- to medium-grained, quartz-cemented sandstone characterized by parallel lamination, low-angle cross-bedding, ripple- and herringbone cross-bedding. In some degree Ophiomorpha sp. traces and enrichments of rip-up clay clasts are present. Lithotype 3A is grey, fine-grained sandstone mainly characterized by large-scale low-angle crossbedding, hummocky- and swaley cross-bedding as well as parallel lamination. Occasional obliteration by biogenic activity mainly Ophiomorpha sp. and Chondrites sp. burrows. Partly oblique and vertical escape traces observable. Lithotype 3B is grey, strongly bioturbated, shaly silt- to fine-grained sandstone. Determinable burrow structures comprise Ophiomorpha sp., Ophiomorpha nodosa, Planolites sp., Chondrites sp., Helminthopsis sp. and Palaeophycus sp. In isolated cases, biogenically distributed medium- to coarse sand grains discernable. Very rare relict primary sedimentary structures visible. Lithotype 4A is interbedded black shale and thin silt- to fine-grained sandstone layers and lenses. Bedding types correspond to lenticular and wavy bedding. Quite frequent occurrence of pyrite nodules and occasional Planolites sp. and Chondrites sp. trace fossils.

The study focuses on using machine learning methods to forecast lithology facies by analyzing well log data from the Miano-5 site using RandomForest Classification and GradientBoosting Classification. GradientBoosting showed an accuracy rate of 85.7%, whereas RandomForest showed a slightly higher rate of 87.1%. This

discrepancy in accuracy raises the likelihood that RandomForest was more effective at recognising the underlying patterns in the data, which led to more accurate lithology facies predictions.

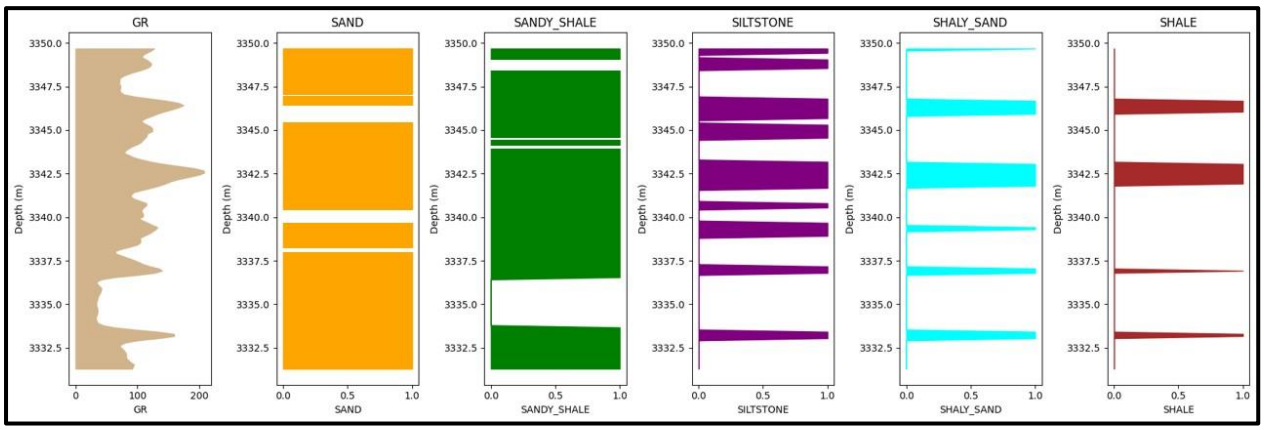


Figure 8 Facies predicted on Miano-5 using Random Forest Classification model

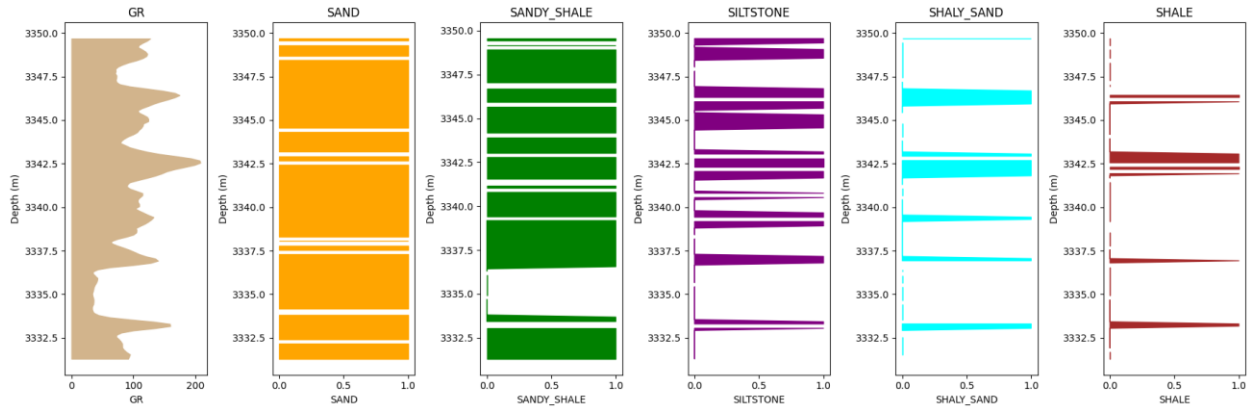
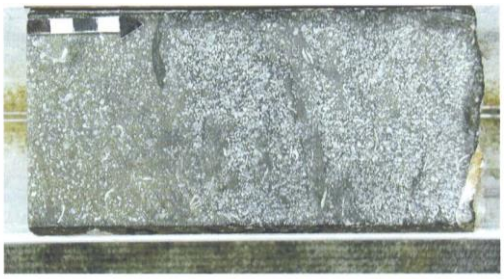


Figure 9 Facies predicted on Miano-5 using Gradient Boosting Classification model

Interesting insights were obtained from looking more closely at the accuracy rates for each lithology column. Across a range of lithology types, RandomForest attained accuracy values between 87.16% and 92%; sand (92%) and sandy shales (88%), in particular, showed especially high accuracy values. These findings show how well RandomForest classified these lithologies; this efficacy may be related to its ability to manage complex interactions within the data and identify small patterns specific to each type of lithology. GradientBoosting, on the other hand, demonstrated less accuracy across lithology columns, ranging from 86% to 91%. These results imply that GradientBoosting may have had slightly more difficulty capturing the finer properties of few lithology types. This might have led to lower accuracy rates. GradientBoosting was still able to attain good accuracy.

Furthermore, with an impressive total accuracy of 92.31%, the use of MLP-SVM for predicting facies using RGB log data showed encouraging results. This high accuracy rate suggests that the lithology facies predictions made by MLP-SVM were based on successful training from the RGB log data. The robustness of the model's performance is further demonstrated by the accuracy, recall, and F1 score metrics of 0.96, 0.93, and 0.94, respectively. These metrics show that the model can achieve high precision while simultaneously catching a large proportion of relevant occurrences.

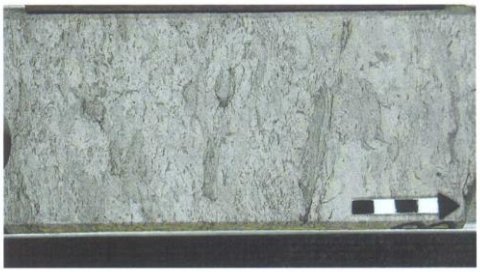
Depth: 3316.0-3316.15 m Lithotype: 1A Facies: Transgressive offshore facies



Depth: 3316.38-3316.58 m reworked shelfal facies Lithotype: 1A, 1B Facies: Transgressive offshore/

Depth: 3320.35-3320.48 m Lithotype: 2A Facies: Tidal channel-fill facies

Depth: 3321.10-3321.21 m Lithotype: 2A Facies: Tidal channel-fill facies



Depth: 3346.48-3346.70 m dominated lower shoreface Lithotype: 3B Facies: Weakly to moderately storm-

Depth: 3335.44-3335.65 m lower shoreface Lithotype: 3A, 3B Facies: Moderately storm-dominated

Depth: 3333.48-3333.71 m lower- to middle shoreface Lithotype: 4A, 3C, 3D Facies: Strongly storm-dominated

Furthermore, the analysis of the confusion matrix provided helpful details about the misclassifications the models created. As an example, two examples of class 5 were incorrectly classed as class 4, whereas one case of class 6 was wrongly classified as class 2. These misclassifications point to possible areas where the predicted accuracy of the models might be improved, such improving feature selection or modifying model hyperparameters to more effectively discriminate across similar lithology types.

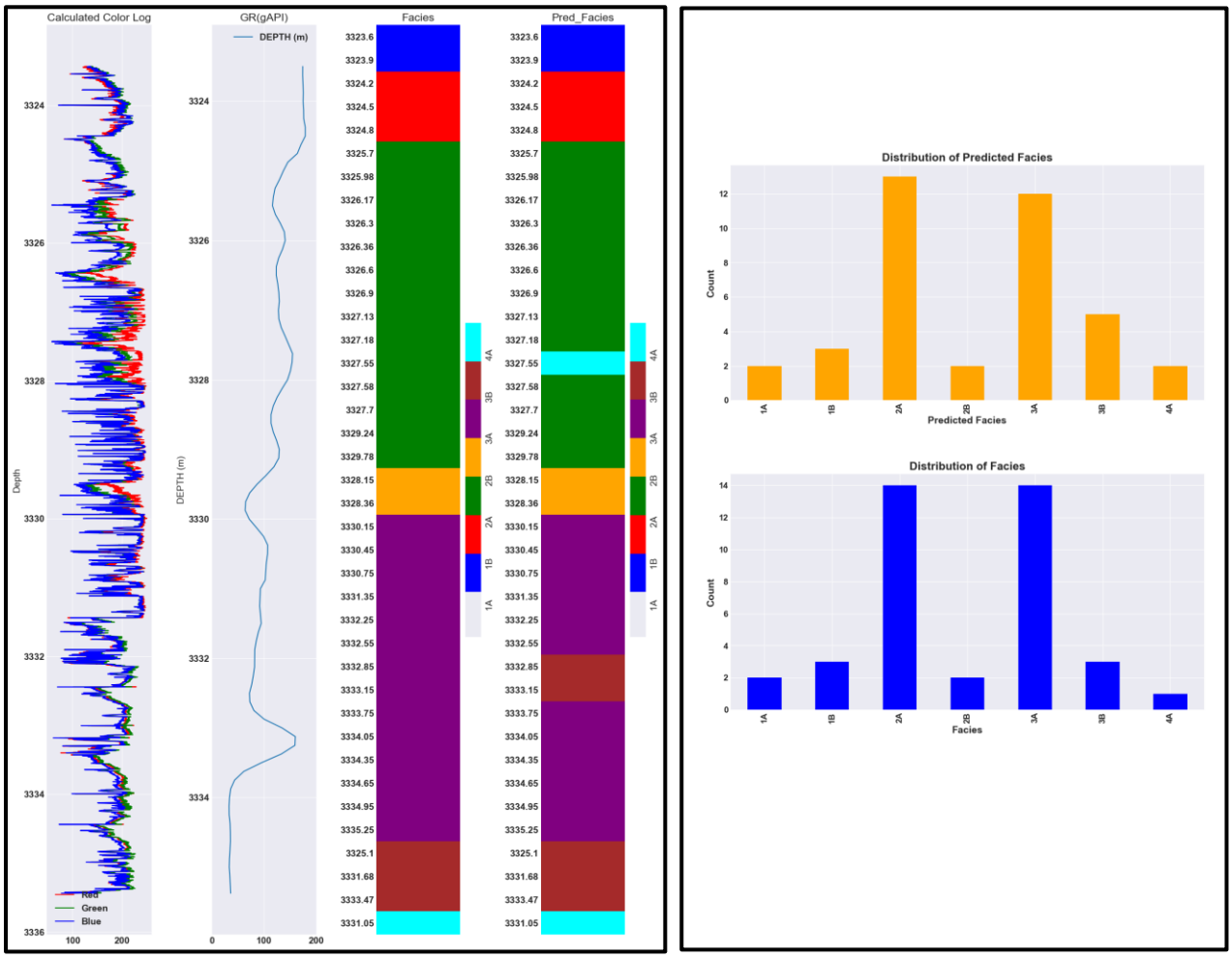


Figure 10 Facie predicted on Miano 5 using RGB log along with actual facies (left). Comparison of distribution of actual and predicted facies (right)

	1A	1B	2A	2B	3A	3B	4A
1A	2	0	0	0	0	0	0
1B	0	3	0	0	0	0	0
2A	0	0	13	0	0	0	1
2B	0	0	0	2	0	0	0
3A	0	0	0	0	12	2	0
3B	0	0	0	0	0	3	0
4A	0	0	0	0	0	0	1

Fig 11 Confusion Matrix of MLP SVM model prediction

Conclusion

The study uses borehole data of Miano block in lower indus basin to analyze the application of automatic lithology prediction. Lithology of core from 3 wells were labelled at finer scale, classifying into 7 classes. RGB channels

summarize the image at each depth interval, with model accuracy >92% for clastic lithology classes. RGB data was faster to compute and had more accurate results, likely because RGB data summarizes data important for lithology prediction.

The well log data model showed some limitations in predictive accuracy, although the core images performed well. The Random Forest and Gradient Boosting models achieved accuracies of 87.1% and 85.7%, respectively. This research demonstrates the potential of using standard consumer desktop equipment, freely available data, and software to convert large amounts of previously imaged core into a standardized digital format. This format is enriched with specific geological insights and interpretations, allowing geoscientists to better integrate extensive subsurface datasets and enhance core-image data characterization. This approach benefits sectors such as mining, hydrogeology, geothermal energy, carbon capture, and geotechnical research. Moreover, the workflow proves adaptable for basin-wide or other large-scale subsurface studies requiring lithology identification.

Although geoscientists' interpretations will always be necessary, machine-learning processes can supplement and expedite certain repetitious activities (such core description), particularly when used at the large, basin scale. An expert geoscientist's physical examination of the core material will never be replaced by the interpretation of core photos alone.

References

1. Ahmad, N., Fink, P., Sturrock, S., Mahmood, T. and Ibrahim, M., 2012. Sequence stratigraphy as predictive tool in lower Goru fairway, lower and middle Indus platform, Pakistan. *PAPG, ATC*, *1*, pp.85-104.
2. Baldwin, J.L., Bateman, R.M. and Wheatley, C.L., 1990. Application of a neural network to the problem of mineral identification from well logs. The Log Analyst, 31(05).
3. Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, pp.785–794. doi:https://doi.org/10.1145/2939672.2939785.
4. Dunham, M. W., Malcolm, A., and Kim Welford, J. (2020). Improved Well-Log Classification Using Semisupervised Label Propagation and Self-Training, with Comparisons to Popular Supervised Algorithms. Geophysics 85 (1). doi:10.1190/geo2019-0238.1O1–O15.
5. Fu, D., Su, C., Wang, W. and Yuan, R., 2022. Deep learning based lithology classification of drill core images. Plos one, 17(7), p.e0270826.
6. Hall, B., 2016. Enthought,". Facies classification with machine learning", SEG: The Leading Edge, 35, pp.906-909.
7. Hall, M. and Hall, B., 2017. Distributed collaborative prediction: Results of the machine learning contest. The Leading Edge, 36(3), pp.267-269.
8. Liu, L., Chen, J., Fieguth, P., Zhao, G., Chellappa, R. and Pietikäinen, M. (2018). From BoW to CNN: Two Decades of Texture Representation for Texture Classification. International Journal of Computer Vision, 127(1), pp.74–109. doi:https://doi.org/10.1007/s11263-018-1125-z.
9. Luthi, S. M. (1994). Textural segmentation of digital rock images into bedding units using texture energy and cluster labels. *Mathematical Geology, 26*(2), 181-196. https://doi.org/10.1007/bf02082762
10. Meyer, R., Martin, T., & Jobe, Z. (2020). CoreBreakout: Subsurface core images to depth-registered datasets. *Journal of Open Source Software, 5*(50), 1969. https://doi.org/10.21105/joss.01969
11. Passey, Q.R., Dahlberg, K.E., Sullivan, K.B., Yin, H., Brackett, R.A., Xiao, Y.H. and Guzmán-Garcia, A.G., 2006. AAPG Archie Series, No. 1, Chapter 1: The Clastic Thin-bed Problem.
12. Quinlan, J.R. (1986). Induction of decision trees. Machine Learning, [online] 1(1), pp.81–106. doi:https://doi.org/10.1007/bf00116251.
13. Tucker, M.E. (2011). Sedimentary Rocks in the Field. John Wiley & Sons.
14. Wolf, M. and Pelissier-Combescure, J. (1982b). Faciolog - automatic electrofacies determination.
15. Zhang, L.F., Pan, M. and Li, Z.L., 2020. 3D modeling of deepwater turbidite lobes: a review of the research status and progress. Petroleum Science, 17, pp.317-333.