

Enhancing lithofacies machine learning predictions with gamma-ray attributes for boreholes with limited diversity of recorded well logs



David A. Wood

DWA Energy Limited, Lincoln, United Kingdom

ARTICLE INFO

Keywords:

Rolling average derivatives

Log-curve volatility

Lithofacies log characteristics

Confusion analysis

Gamma-ray attributes

Well-log feature augmentation.

ABSTRACT

Derivative and volatility attributes can be usefully calculated from recorded gamma ray (GR) data to enhance lithofacies classification in wellbores penetrating multiple lithologies. Such attributes extract information about the log curve shape that cannot be readily discerned from the recorded well log data. A logged wellbore section for which 8911 data records are available for the three recorded logs (GR, sonic (DT) and bulk density (PB)) is evaluated. That section demonstrates the value of the GR attributes for machine learning (ML) lithofacies predictions. Five feature selection configurations are considered. The 9-var configuration including GR, DT, PB and six GR attributes, and the 7-var configuration of GR and the six GR attributes, provide the most accurate and reproducible lithofacies predictions. The other three feature configurations evaluated do not include the GR attributes but just one to three of the recorded log features. The results of seven ML models and two regression models reveal that K-nearest neighbor (KNN), random forest (RF) and extreme gradient boosting (XGB) are the best performing models. They generate between 14 and 23 misclassification from 8911 data records for the 9-var model. Multi-layer perceptron (MLP) and support vector classification (SVC) do not perform well with the 7-var model which lacks the PB feature displaying the highest correlation with facies class. Annotated confusion matrices reveal that KNN, RF and XGB models can effectively distinguish all facies classes for the 9-var and 7-var configurations (that includes the GR attributes), whereas none of the models can achieve that outcome for the 3-var configuration (that excludes the GR attributes). Accurately distinguishing lithofacies using well-log data in sedimentary sections is an important objective in applied geoscience. The straightforward, GR-attribute method proposed works to improve confidence in ML-lithofacies classifications based on limited recorded well-log data.

1. Introduction

Lithofacies classification is a key requirement of sedimentological interpretation and has been a cornerstone of applied sedimentology and depositional-environment interpretations in the oil and gas industry since the 1970s (Selley, 1978). The possibility of exploiting recorded well-log data to assist in lithofacies classification dates back to the mid-twentieth century (Russell, 1944; Krumbein and Sloss, 1951). At that time, the well logs available that were able to provide reasonable resolution for lithofacies interpretation were limited to gamma ray (GR) and a compressional velocity/acoustic log providing travel time (DT). Nevertheless, since that time those well logs have been widely exploited for lithofacies analysis and stratigraphic sequencing (Scholle and Spearing, 1982; Rogers et al., 1992; Emery and Myers, 1996).

Well-log data has been most successfully exploited to distinguish lithofacies in clastic sedimentary sequences, with the shape and sensitivity of the GR log to sandstone/shale facies sequences well documented

(Rider, 1986, 1990; Cant, 1992). Additionally, the GR responses to grain size of clastic formations (Hurst, 1990) make it possible to distinguish important lithofacies features such as fining- or coarsening upwards (Kesslar and Sachs, 1995; Faga and Oyeneyin, 2000). In carbonates, and other more varied and heterogeneous lithological sequences additional geological and mineralogical information is typically required to complement well-log data sufficiently to provide reliable lithofacies classification (Reverdy et al., 1983; Halotel et al., 2020). Such analysis typically requires core data and specialized well logs capable of distinguishing mineralogy, which are expensive and time-consuming to acquire. Most wells drilled, particularly development wells, do not record comprehensive well-log suites over the entire wellbore sections drilled. Consequently, for most well sections drilled, only a limited set of well logs are available. With such limited well-log data, traditional models that attempt to predict lithofacies classes in multi-lithology sequences tend to lack precision and reliability.

Statistical methods, specifically regression and cluster analysis

E-mail address: dw@dwasolutions.com.

methods (Goncalves et al., 1995), have been applied to well-log data to assist in lithofacies prediction since the 1980s (Busch et al., 1987). However, in multi-lithology sections well-log data typically displays non-linear relationships with lithofacies, making such methods unreliable and relatively inaccurate. Various artificial-neural-network models, particularly multi-layer perceptrons (MLP), are extensively used to predict facies classes from well-log data (Rogers et al., 1992; Puskarczyk, 2019; Hosseini et al., 2020). One attraction of MLP is that for complex carbonate sequences they can generally outperform regression, probabilistic (e.g., Bayesian classification) and fuzzy classifier methods in lithofacies classification tasks using multiple well-log curves (Dubois et al., 2007; Agrawal et al., 2022). Neural networks have also been usefully hybridized with a Markov transition matrix (Feng et al., 2018; Feng, 2020) to form Bayesian neural networks (Feng, 2021) for lithofacies classification using well-log data.

In recent years, the high lithofacies prediction performance of ensemble and boosted-tree-based algorithms has become apparent (Xie et al., 2019). This is particularly the case for carbonate and complex lithological sequences by several studies (Hall, 2016; Hall and Hall, 2017; Al-Mudhafar, 2017; Bestagini et al., 2017, Shashank and Mahapatra, 2018). Adjusting the control parameters of boosted-tree models, such as extreme gradient boosting (XGB) with Bayesian optimization can improve their ensemble-learning performance for lithofacies classification (Sun et al., 2020). Additionally, regression-free K-nearest neighbor (KNN) (Merembayev et al., 2021) and the transparent open box algorithm Wood (2019, 2020) have been successfully applied to challenging lithofacies and stratigraphic classification and data mining using well-log data. Combinations of data-matching algorithms with clustering techniques have also been usefully applied (Potratz et al., 2021). Integrating ML models, clustering techniques (Fadokun et al., 2020), geological data (Ippolito et al., 2021) and bio-stratigraphic information (Tran et al., 2020) with well log data can also be effective for facies classification purposes, in cases where sufficient data is available.

Recently (Wood, 2022), has demonstrated that calculating derivative and volatility attributes from a GR log, associated with a highly variable clastic sequence, is impressively effective at classifying several facies characteristics typically displayed by sandstones and shales, without recourse to other well logs. This method is particular attractive for assessing wellbore sections for which very few well log curves are recorded and/or for which core, mineralogy and biostratigraphy data are lacking. It is the objective of this study to build on that GR method, specifically to demonstrate that it is also highly effective at lithofacies classification in multiple lithology sections with limited well-log coverage. A logged section though varied lithologies (clastic, carbonate, salt and evaporite) is evaluated with this new GR-attribute technique to assess its potential applying seven distinct ML models and two linear regression models. The novelty of this study is that it applies the GR-attribute, lithofacies-classification technique for the first time to a complex sedimentary section composed of multiple lithologies.

2. Method and materials

2.1. Well-log feature augmentation

It has been recently demonstrated using a synthetic GR log that derivative and volatility attributes extracted from the gamma-ray curve could be effectively used to distinguish generic facies characteristics (e.g., fining upwards) in a clastic section (Wood, 2022). That methodology is adapted here for application to lithofacies classification in a logged section from a real wellbore penetrating a diverse sequence of lithologies (i.e. limestone, dolomitic mudstone, halite, evaporite, shale, siltstone and sandstone). Six attributes are calculated from the recorded GR log: three derivative attributes; and, three volatility attributes.

2.1.1. GR derivative attributes

The first derivative ($d'GR$) is calculated with Eq. (1).

$$d'GR = (GR_d - GR_{d-1}) / \text{Abs}(d - (d - 1)) \quad (1)$$

where: GR_d = recorded GR value at GR curve location d (depth in feet or meters); GR_{d-1} = recorded GR value at GR curve location $d-1$.

The second derivative ($d''GR_{dn}$) is calculated with Eq. (2).

$$d''GR_{dn} = (d'GR_d - d'GR_{d-n}) / \text{Abs}(d - (d - n)) \quad (2)$$

where: n = a specified depth interval. An n value of 10 data points, or 1 m, is applied in this study, established via trial and error analysis.

The first derivative's moving-average metric (SMA) is calculated with Eq. (3).

$$SMA d'GR_{dn} = \left(\sum_{i=1}^{i=n} d'GR_{d-i} \right) / n \quad (3)$$

Again, an n value of 10 data points, or 1 m, is applied in this study, established via trial and error analysis. The optimum value of n is likely to depend on the nature of the lithological section drilled. Finely interbedded strata with high volatility are likely to be better characterized by relatively low values of n . On the other hand, massive bedded sequences may be more effectively characterized by higher values of n . Trial and error tests are required to determine the n value, or range of values, most suitable for specific drilled sections.

2.1.2. GR volatility attributes

Wood (2022) proposed the use of volatility attributes as calculated by the financial sector in the analysis of trends of traded market indices.

The instantaneous movement or change in GR signal ($LnGR_i$) represents an intermediate step in establishing volatility. It is calculated with Eq. (4)

$$LnGR_{i(d)} = \ln(GR_d / GR_{d-1}) \quad (4)$$

where: $i(d)$ = interval between depths $d-1$ and d (i.e., between sequential GR recorded points).

Standard deviation is used to calculate volatility (σ_i) of $LnGR_{i(d)}$ for a specific sequence of GR data points. It is calculated with Eq. (5).

$$\sigma_{i(dn)} = \sqrt{\frac{\sum_{i=1}^n (L_{d-i} - L_{d-1} \text{mean})^2}{n - 1}} \quad (5)$$

where: $\sigma_{i(dn)}$ = interval volatility between depths $d-1$ and $d-n$; L_{d-i} = $LnGR_{i(d)}$ for each GR data point in the interval between depths $d-1$ and $d-n$; $L_{d-1} \text{mean}$ = arithmetic mean of $LnGR_{i(d)}$ between depths $d-1$ and $d-n$; n = interval over which σ_i is calculated. An n value of 10 data points, or 1 m, is applied in this study, established via trial and error analysis.

The third volatility attribute is a simple moving average (SMA σ_i) of the GR volatility attribute. It is calculated with Eq. (6).

$$SMA\sigma_{i(dn)} = \left(\sum_{i=1}^{i=n} \sigma_{d-i} \right) / n \quad (6)$$

Again, an n value of 10 data points, or 1 m, is applied in this study, established via trial and error analysis. As for the derivative SMA (Eq. (3)), the n value can be varied to suit the drilled geological section under consideration. Furthermore, the value of n could also be different for Eqs. (5) and (6), if the trial and error analysis suggests it.

The three derivative GR attributes and the three volatility GR attributes defined by Eqs. (1)–(6) are used in this study as additional log features to complement the recorded GR data points. Each of these

attributes emphasizes subtle changes in the GR trend as it changes with depth, and these subtleties can be usefully exploited by machine learning algorithms to classify lithofacies more precisely.

2.2. Borehole logged section evaluated

The logged wellbore section evaluated in this study is from the Triassic section penetrated in the Winterborne Kingston –1 (WK-1) well bore drilled in 1981 in the western part of the petrolierous Wessex Basin in the Southern United Kingdom (Rhys et al., 1982). The section of interest represents an interval of about 900 m between measured depths 1560–2460 m. It is selected for evaluating the potential benefits of GR attributes for multiple lithology facies prediction for several reasons:

- The section includes a substantial range of primarily non-clastic lithologies in a relatively small depth interval;
- Some of the facies involved display quite similar absolute GR ranges and patterns;
- Substantial parts of the section involve thinly interbedded lithologies associated with quite rapidly and extensively oscillating GR signals. This makes some of the lithofacies quite difficult to distinguish by machine learning methods and/or cluster analysis using the GR log in combination with the DT and PB logs, as previously noted (Newell et al., 2021).

These feature of the WK-1 Triassic section pose a meaningful challenge for testing GR attributes ability to improve lithofacies prediction in multi-lithology sections, i.e., quite distinct from the clastic (sand/shale) sequences to which the GR attribute method has recently been successfully developed and applied (Wood, 2022).

The Triassic stratigraphy of the Western Wessex basin is summarized in Fig. 1.

The lowest stratigraphic group of the well-log sequence evaluated is the fining upwards clastic sequence of the Sherwood Sandstone Group, resting unconformably on the shales of the Aylesbeare Mudstone Group. The Sherwood Sandstone, specifically the Otter Sandstone Formation (locally) is overlain by the variable lithologies of the Mercia Mudstone Group. This “mudstone” group contains abundant dolomitic siltstone interbeds above and below the central evaporitic sequence constituting

the Dorset/Somerset Halite. Above the halite zones, within the Branscombe Mudstone Formation, a thin but distinctive (from its well-log characteristics) anhydrite and limestone zone exists at about 1865 m depth. This is a representation of the Red Rock Gypsum Layer distributed sporadically in the Wessex Basin (Hounslow and Ruffell, 2006). The Blue Anchor formation is the uppermost formation of the Mercia Mudstone Group. Its grey/green colored dolomitic siltstones with sporadic anhydrite/gypsum layers display well log characteristics (Howard et al., 2008), in terms of GR, DT and PB values, which are difficult to distinguish from those of the Otter sandstones. The uppermost stratigraphic group of the well-log sequence evaluated is the cream/grey colored, fine grained Lilstock limestone of the Penarth Group (also referred to as the White Lias). That limestone represents the uppermost Triassic formation (Rhaetian age) and it overlain by shales of the Blue Lias (Jurassic).

The GR, PB and DT, well log curves of the ~900 m thick section of interest are sampled at 10 cm intervals (10 data records/meter) with five lithofacies classes distinguished. Those lithofacies are allocated to classes 1 to 5:

- 1 Limestone with sparse anhydrite
- 2 Dolomitic mudstones/siltstones
- 3 Shale
- 4 Salt
- 5 Shaly sandstones and siltstones

It is the lithofacies class numbers 1 to 5 that are predicted by the ML models evaluated.

Fig. 2 displays the recorded well logs as sampled from recorded curves versus depth. The intervals of limestone, anhydrite, salt/evaporite and shale, belonging to lithofacies 1, 4 and 3, display quite distinctive GR, PB and DT signals and that can be readily distinguished by most ML algorithms. However, there is a good deal of overlap in the GR, PB and DT values between the dolomitic mudstones/siltstones and shaly sandstones and siltstones lithofacies (2 and 5) of the Mercia Mudstone and Sherwood Sandstone formations. Consequently, the ML algorithms struggle to distinguish between portions of lithofacies 2 and 5.

There are intrinsic transitional zones between each lithofacies (Fig. 2), which are more gradational between some intervals (e.g., the top of the Sherwood Sandstone and base of the Mercia Mudstone) but quite

Stratigraphic Column for Triassic Section of the Western Wessex Basin					
Era	Epoch	Group	Formation	Member	Lithology
Triassic	Jurassic	Hettangian	Blue Lias		
	Rhaetian	Penarth	Lilstock	Longport Limestone	Limestones
			Westbury		Unconformity
	Norian	Mercia Mudstone	Blue Anchor	Upper Red Marl	Mudstones / Dolomitic Siltstones
			Branscombe Mudstone		Shales / Limestone bands
	Carnian		Dorset/Somerset Halite	Evaporite Series	Halite /Evaporites
	Ladinian		Sidmouth Mudstone	Lower Red Marl	Mudstones / Dolomitic Siltstones
	Anisan	Sherwood Sandstone	Otter Sandstone		Shaly / Silty Sandstones
			“BSPB”		Unconformity
	Scythian	Aylesbeare Mudstone	Littleham Mudstone		Shales
Permian	Lopingian				

Fig. 1. Summary stratigraphy and lithology of the Wessex Basin Triassic section relevant to the Winterborne Kingston WK-1 wellbore location. Incorporating information from Underhill and Stoneley (1998), Hounslow and Ruffell (2006); Howard et al. (2008); Kaya (2015). BSPB refers to the Budleigh Salterton Pebble Beds.

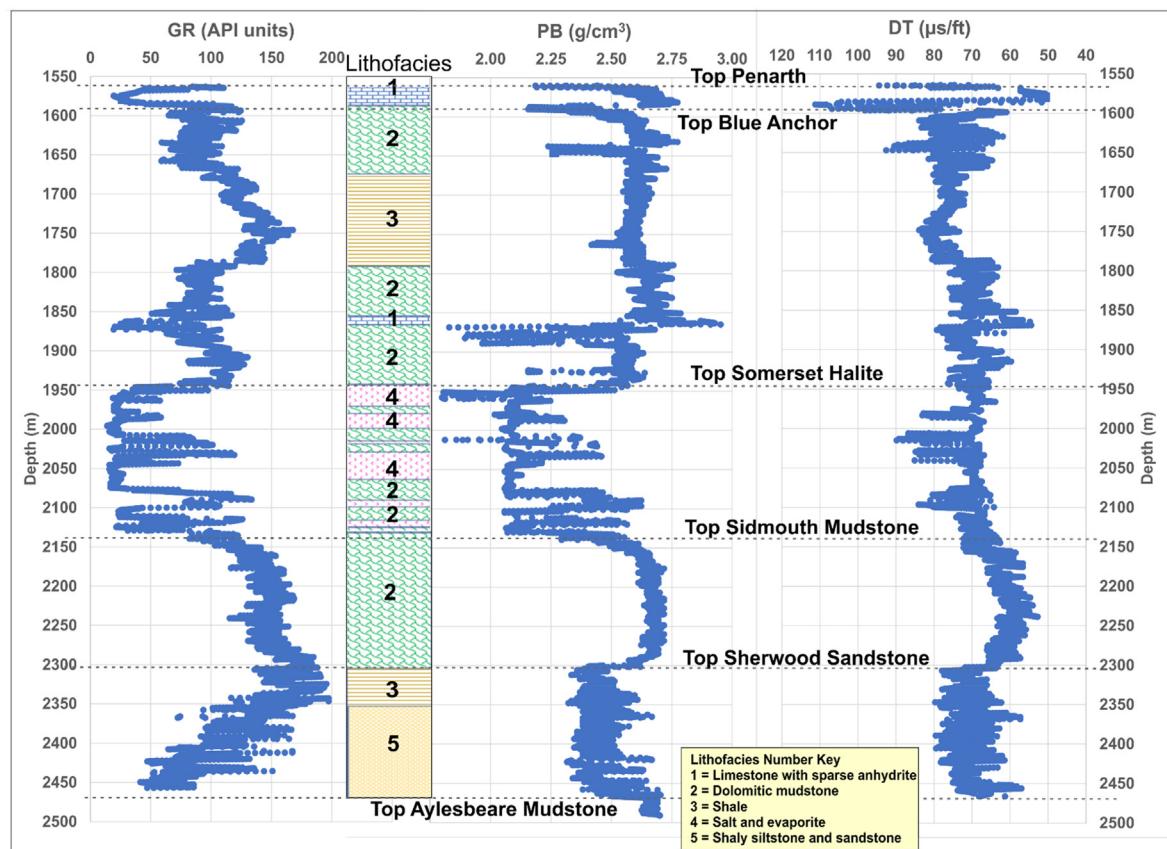


Fig. 2. GR, PB and DT well logs recorded for the Triassic section of the WK-1 wellbore with 5 distinct lithofacies sequences distinguished as classes 1 to 5.

abrupt between other intervals (e.g., the contact at the top of the Somerset Halite with the overlying dolomitic mudstone). One of the advantages of extracting and using the GR-attributes from the GR recorded data is that they are better able to characterize the gradational transition zones between lithofacies than the recorded GR data on its own.

Figs. 3 and 4 display the three GR derivative and three GR volatility attributes, respectively, calculated for each data record versus depth. Notice that the anhydrite/limestone band within the Branscombe Mudstone Formation stands out as extreme values in all of the calculated derivative/volatility attributes. Also, these attributes show quite distinctive distributions for the Blue Anchor Formation (near the top of the section) and the Otter sandstones (at the base of the section). Such a distinction cannot be discerned from the GR data in isolation.

Table 1 compares the distribution statistics for the nine well log features under consideration (three recorded well logs GR DT and PB; six calculated GR attributes). The cumulative probability distributions of these variables and the lithofacies number are displayed in Fig. 5. The derivative attributes, in particular, display quite expansive ranges of values. In order to calculate the moving average attributes, the first few data points of the interval are excluded from the data record sequence evaluated. The sequence evaluated (Table 1) involves a continuous sequence of 8911 data records (1565.9 m–2456.9 m inclusive). About 53% of those data records belong to facies class 2 (dolomitic mudstone) distributed through the post-Otter Sandstone sequence.

Fig. 6 displays a heat map of the calculated correlation coefficient values between the log variables and facies class. The values for all columns in Fig. 6 except the last column to the right are for the Pearson's correlation coefficient (R), which assumes that the variable distributions are parametric and essentially based on linear relationships. The right-side column (Fig. 6) displays Spearman's rank correlation coefficients (p), a non-parametric statistic that is more representative of distributions

of variables that follow non-linear relationships. The R values between GR and its attributes are low (<0.2 , left-side column, Fig. 6). The R values between the GR attributes and facies class are substantially lower than those displayed by the GR, DT, PB recorded log data. PB displays the highest negative R value with facies class (-0.49), compared to -0.21 for GR and $+0.21$ for DT.

The p values for GR (-0.07), DT ($+0.26$) and PB (-0.54) with facies class are quite distinctive from the R values. This difference is indicative of non-linearity influencing the relationships between the three measured log curves and facies class, making the p values more representative of those distribution relationships than the R values. The GR attributes are individually poorly correlated in terms of both R and p values (<0.15) with facies class. Such relationships suggest that regression-type models, particularly those based on linear assumptions are unlikely to be able to predict facies class accurately using the recorded well log variables and/or, in particular the calculated GR-attribute variables.

2.3. Regression and machine-learning algorithms applied

Two linear regression algorithms, together with seven algorithms configured for classification applying distinct ML methods are used to model the log data recorded from the Triassic section of well WK-1, together with the six calculated GR attributes. The algorithms are coded in Python and developed around SciKit Learn functions (SciKit Learn, 2022a).

The algorithms evaluated have distinctive methodologies and can be categorized as regression-based (OLR, SGD and SVC), tree/ensemble-based (ADA, DT, RF, XGB), data matching (KNN) and neural network (MLP). These nine models are now defined, in alphabetical order, with the first cited reference for each method referring to the original

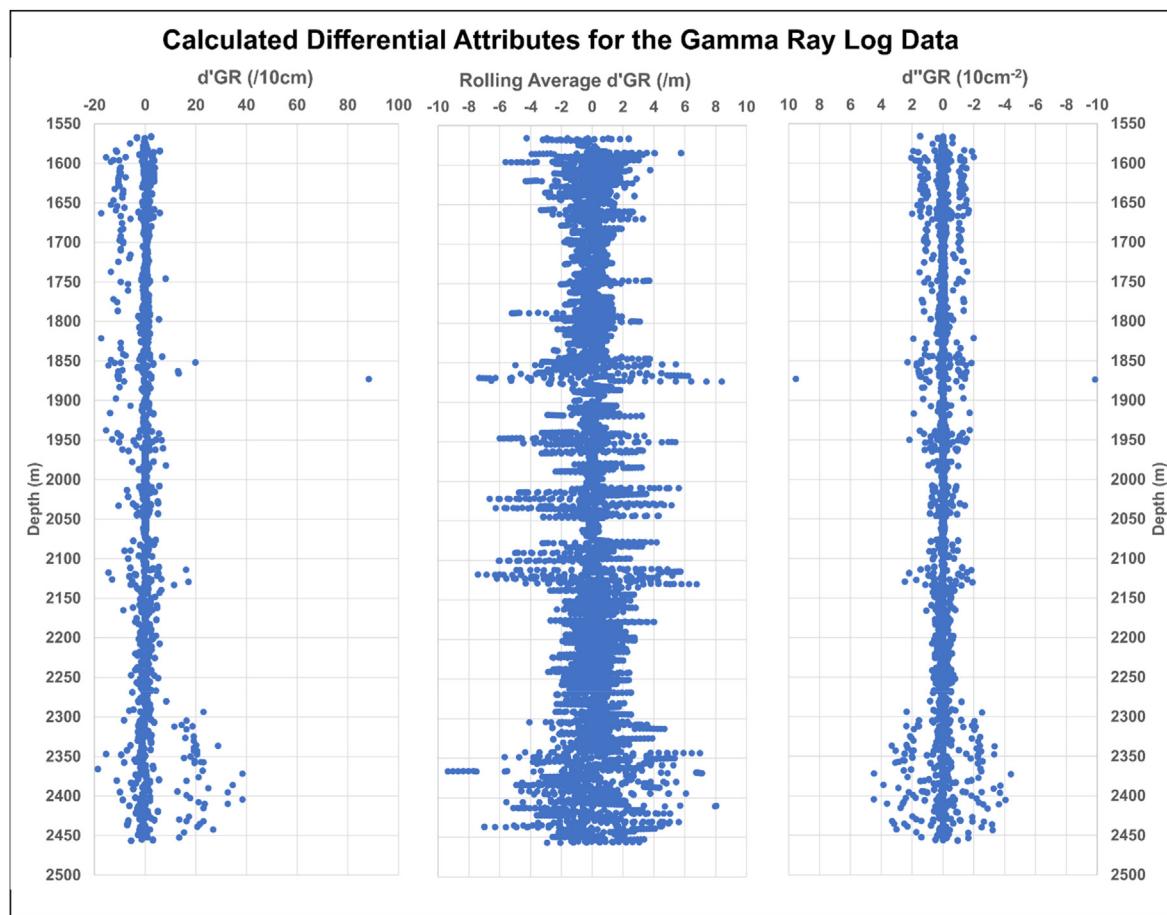


Fig. 3. Three GR derivative attributes calculated for each GR data point of the Triassic section of the WK-1 wellbore. These derivative attributes are the first derivative ($d'GR$), the moving average per meter of the first derivative (SMA $d'GR$) and the second derivative ($d''GR$).

developers of the algorithm for ML applications, and the subsequent citations referring to example recent applications of each model to address lithofacies classification problems.

ADA: Adaboost (Freund and Schapire, 1997; Wrona et al., 2018).

DT: Decision tree (Quinlan, 1986; Sarkar and Majundar, 2020).

KNN: K-nearest neighbor (Fix and Hodges, 1951; Merembayev et al., 2021).

MLP: Multi-layer perceptron (Rosenblatt, 1958; Hosseini et al., 2020).

OLR: Ordinary least-squares regression (Goldberger, 1964; Gao, 2011).

RF: Random forest (Ho, 1998; Kim et al., 2018; Farzi and Bolandi, 2016).

SGD: Stochastic gradient descent multi-linear regression (Bottou, 1998; Agrawal et al., 2022).

SVC: Support vector classifier (Cortes and Vapnik, 1995; Mandal and Rezaee, 2019).

XGB: Extreme gradient boosting (Chen and Guestrin, 2016; Bestagini et al., 2017).

All the algorithms evaluated have been extensively deployed for lithofacies modelling and prediction. These algorithms are well described in the literature (see for example citations associated with the models in the list above), therefore their detailed methodologies are not described here. However, the models do require configuration and tuning adjustments to suit each dataset. This is achieved by establishing optimum values for the control/hyperparameters.

Model architecture and hyperparameter selection involves several techniques. Trial and error is important to establish a range of suitable structures and control values. Further refinement of the trial-and-error results can be achieved rapidly using grid search (SciKit Learn, 2021a)

and Bayesian optimization (SciKit Learn, 2021b) to rapidly assess a wide range of control values within specified ranges. These techniques make it possible to locate the optimum control parameter values for each model for the dataset under investigation, which are shown for the models used in this study in Table 2. Applying the optimum configurations and control parameters, the models can then be evaluated with K-fold cross validation (SciKit Learn, 2022b) to establish their repeatability with various random splits of the dataset into training and testing subsets.

2.4. Data preprocessing

The well log data and calculated GR attributes fed as input to the models are each normalized to scale range of -1 to $+1$. Such normalization removes the possibility of scale biases causing certain variables to exert greater influence than others in the facies class predictions. That normalization is conducted with Eq. (7).

$$Normx_i^m = 2 * \left(\frac{x_i^m - xmin^m}{xmax^m - xmin^m} \right) - 1 \quad (7)$$

where: $Normx_i^m$ = variable value adjusted to a -1 to $+1$ scale; x_i^m = actual recorded/calculated value for i th data point in the m th variables distribution; $xmin^m$ = minimum value of m th variable distribution; $xmax^m$ = maximum value of m th variable distribution.

2.5. Metrics for assessing classification performance and misclassification

Errors in lithofacies classification are assessed in this study using several widely used performance measures for which definitions and

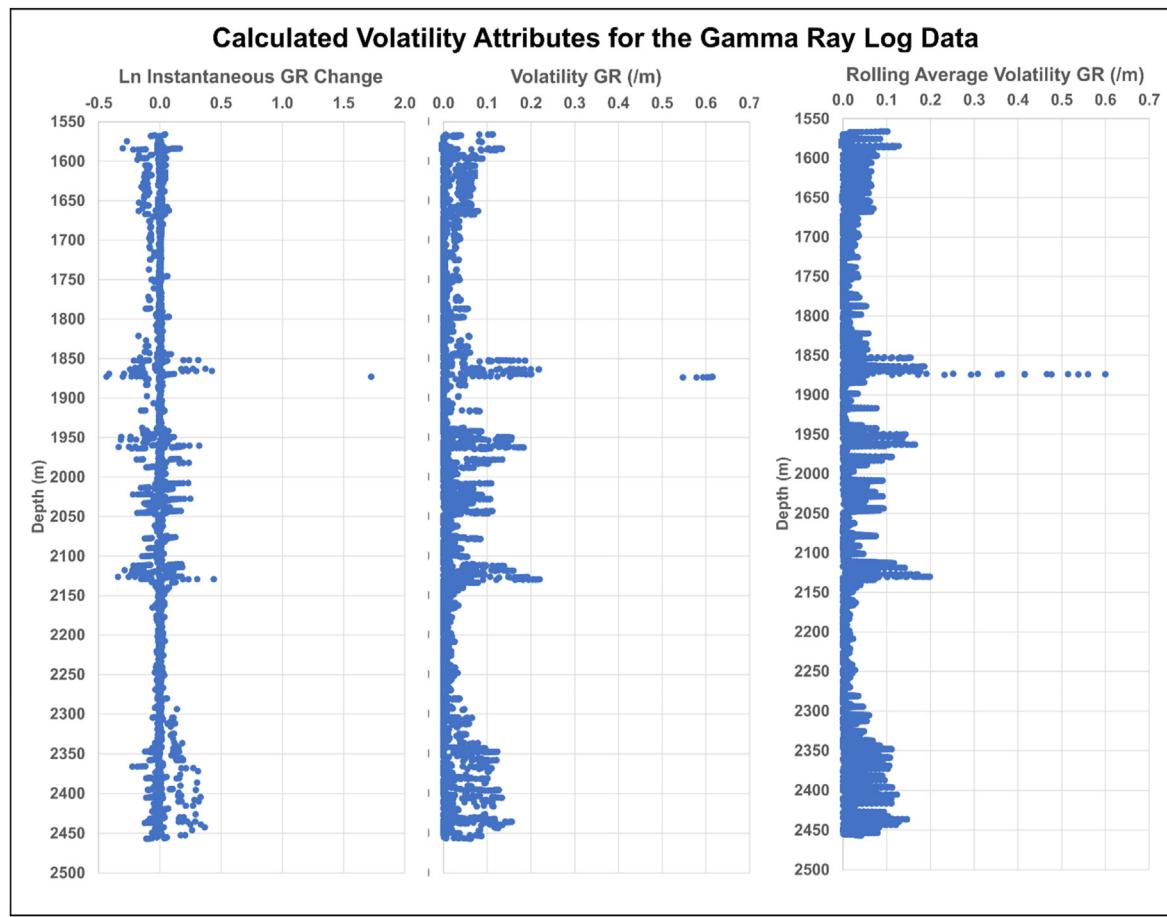


Fig. 4. Three GR volatility attributes calculated for each GR data point of the Triassic section of the WK-1 wellbore. These volatility attributes are the instantaneous change (LnGR_i), volatility (σGR) and the moving average per meter of the volatility ($\text{SMA}\sigma\text{GR}$).

Table 1

Statistical characterization of well log variable and facies class distributions for Triassic well log interval of well WK-1.

Statistical Summary of Recorded Well Log and Gamma Ray Attribute Variables										
8911 Data Points	GR	DT	PB	d'GR	SMA σGR	d"GRdn	LnGRi(d)	$\sigma_i(\text{dn})$	SMA $\sigma_i(\text{dn})$	Facies< Class
Minimum	13.8	50.7	1.80	-18.617	-9.362	-9.863	-0.435	0.00000	0.00000	1
5th Percentile	21.4	58.3	2.08	-4.071	-2.504	-0.803	-0.048	0.00000	0.00007	2
10th Percentile	26.7	60.5	2.10	-2.217	-1.694	-0.405	-0.024	0.00002	0.00038	2
25th Percentile	74.8	66.1	2.39	-0.708	-0.658	-0.082	-0.008	0.00021	0.00228	2
50th Percentile	105.6	69.9	2.56	-0.054	-0.032	0.000	-0.001	0.00541	0.00830	2
75th Percentile	138.1	74.6	2.63	0.844	0.676	0.083	0.009	0.02148	0.02349	4
90th Percentile	153.7	78.3	2.68	2.058	1.653	0.389	0.022	0.05865	0.05156	5
95th Percentile	165.9	80.3	2.69	3.227	2.496	0.701	0.042	0.08624	0.07546	5
Maximum	197.3	111.2	2.95	88.282	8.396	9.533	1.726	0.61501	0.59969	5
Average	101.7	70.1	2.48	-0.001	0.001	0.000	0.000	0.01910	0.01912	2.8
Standard Deviation	44.3	7.0	0.20	3.354	1.552	0.540	0.044	0.03617	0.03085	1.1
Standard Error	0.4689	0.0745	0.0022	0.0355	0.0164	0.0057	0.0005	0.0004	0.0003	0.0117
Coefficient of Variation	0.435	0.100	0.082	-4002	2545	3425	-2669	1.894	1.614	0.397

calculation formulas are provided in Appendix A. These metrics are:

Root Mean Squared Error (RMSE)

Mean Absolute Error (MAE)

Coefficient of Determination (R^2)

Total number of prediction errors ($\sum \text{Error}$)

Accuracy (A)

Precision (P)

Recall (R)

Balanced F-score (F_1)

It is essential to consider the statistical measures of prediction performance ($\sum \text{Error}$, RMSE, MAE and R^2) in conjunction with measures relating specifically to misclassifications (A, P, R, F_1).

3. Results

3.1. Models evaluated for WK-1 Triassic well log data

The base case feature configuration for facies class prediction in the WK-1 Triassic section is the 3-variable model (3-var) considering just the

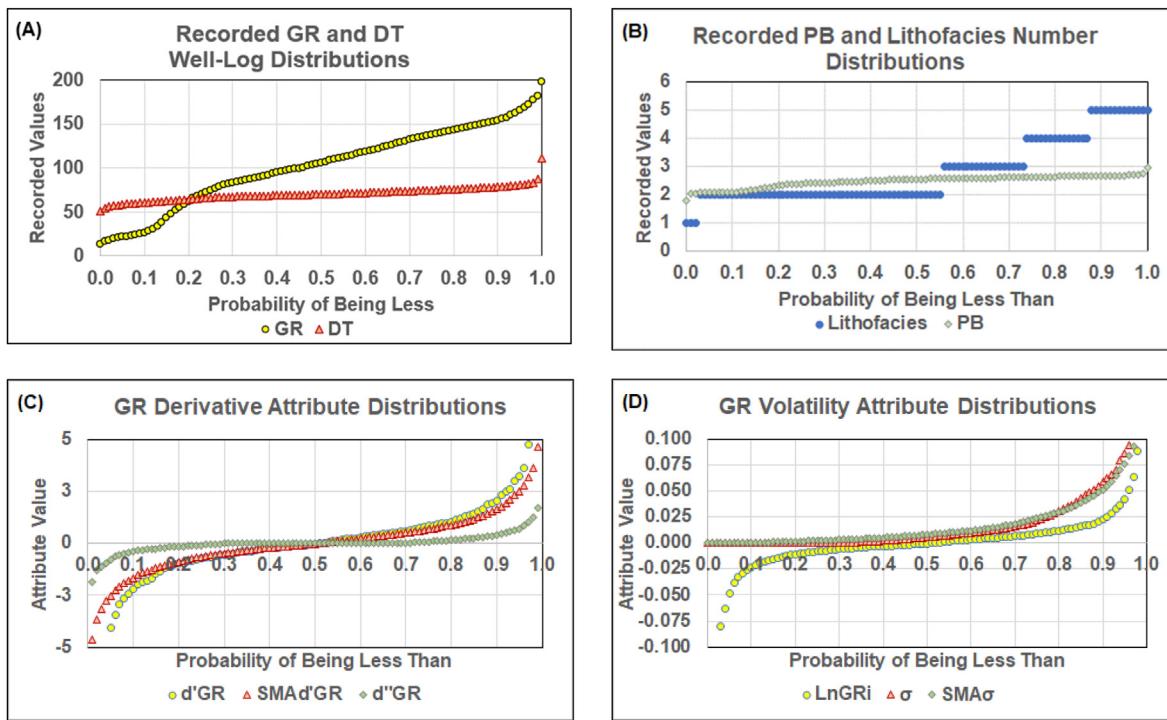


Fig. 5. Cumulative probability distributions for: (A) recorded GR and DT logs; (B) recorded PB log and lithofacies number; (C) calculated GR derivative attributes; and, (D) calculated GR volatility attributes. The extreme ends of some of the GR derivative and GR attribute distributions are excluded to facilitate displays on meaningful scales.

Correlation Matrix (Heat Map) Between Log Curves/ Attributes and Facies Class											
	GR	DT	PB	d'GR	SMA'd'GR	d''GR _{dn}	LnGR _{i(d)}	$\sigma_{i(dn)}$	SMA $\sigma_{i(dn)}$	Facies Class R	Facies Class p
GR	1	-0.0491	0.5999	0.0382	0.1749	-0.0239	0.0353	-0.1527	-0.1690	-0.2106	-0.0725
DT	-0.0491	1	-0.2123	0.0006	0.0436	-0.0286	0.0098	0.0453	0.0618	0.2064	0.2622
PB	0.5999	-0.2123	1	0.0093	0.0307	-0.0039	0.0091	-0.1317	-0.1568	-0.4917	-0.5423
d'GR	0.0382	0.0006	0.0093	1	0.3140	0.7226	0.8852	0.1043	0.0316	-0.0155	-0.1016
SMA'd'GR	0.1749	0.0436	0.0307	0.3140	1	-0.0010	0.3004	0.0811	0.1100	-0.0392	-0.0382
d''GR _{dn}	-0.0239	-0.0286	-0.0039	0.7226	-0.0010	1	0.6427	0.0859	-0.0070	0.0023	-0.0030
LnGR _{i(d)}	0.0353	0.0098	0.0091	0.8852	0.3004	0.6427	1	0.1388	0.0412	-0.0149	-0.1007
$\sigma_{i(dn)}$	-0.1527	0.0453	-0.1317	0.1043	0.0811	0.0859	0.1388	1	0.7214	0.1131	0.0965
SMA $\sigma_{i(dn)}$	-0.1690	0.0618	-0.1568	0.0316	0.1100	-0.0070	0.0412	0.7214	1	0.1295	0.1467
Facies	-0.2106	0.2064	-0.4917	-0.0155	-0.0392	0.0023	-0.0149	0.1131	0.1295	1	1

Note: R refers to Pearson's correlation coefficient (all columns except right-side one); p refers to Spearman's correlation coefficient (right-side column only).

Fig. 6. Pearson's Correlation coefficients for variables associated with the Triassic well log interval of well WK-1. Both Pearson's and Spearman's coefficients are displayed for correlations of the variables with facies class.

recorded GR, PB and DT log data. Nine models are trained and tested with this configuration: two linear models (OLR and SDR) and seven ML classification models (ADA, DT, KNN, MLP, RF, SVC and XGB). The facies prediction results for the training and testing subsets for the 3-var configuration for all model are shown in Table 3.

The perfect results for the training subset achieved by the ADA, DT, KNN and RF models and substantially substantial prediction errors for the testing subsets suggest that those models are tending to overfit training subsets. However, in terms of the classification accuracies achieved by the trained models with the unseen testing subsets, the RF, SVC and KNN models substantially outperform the other models considering the low RMSE, MAE and \sum Error values those models achieve. The linear models generate the poorest facies prediction performance for the 3-var testing subset based on those error metrics. In terms of computer execution time, the XGB and MLP models take substantially longer to execute than the other models. Of the high-performing models, KNN takes the shortest execution time (<6 s) for the 3-var configuration.

In addition to the 3-var base case evaluation four other variable

configuration are evaluated with each of the nine models. The configurations evaluated and compared are:

- 9-var: (GR) (DT) (PB) (d'GR) (SMA'd'GR) (d''GR) (LnGR_i) (σ GR_i) (SMA $\sigma_{i(dn)}$)
- 7-var: (GR) (d'GR) (SMA'd'GR) (d''GR) (LnGR_i) (σ GR_i) (SMA $\sigma_{i(dn)}$)
- 3-var: (GR) (DT) (PB) [Base Case]
- 2-var: (GR) (PB)
- 1-var: (GR)

Two other 2-var configurations were also evaluated (GR with DT, and DT with PB). However, the GR with PB combination delivered better predictions than those other configurations, so only the GR with PB 2-var configuration results are presented.

The facies class prediction performances of the five-feature configurations defined are benchmarked against the 3-var configuration (base case) using each of the nine prediction models considered.

Table 2

Set up and control parameters for regression and ML algorithms applied to predict lithofacies class for the WK-1 Triassic well log section evaluated.

Algorithm	Control Parameter Values Applied
Adaboost (ADA)	Number of estimators = 750; learning rate = 0.01; base estimator is DT with depth = 250; splitter = best
Decision Tree (DT)	Maximum depth = 5000; splitter = best; splitting criteria = gini
K Nearest Neighbor (KNN)	Weighted using Manhattan distance; number of neighbours assessed (K) K = 5 for 2-var and 3-var models; K = 2 for 7-var and 9-var models; k = 25 for 1-var model
Multi-layer Perceptron (MLP)	3 hidden layers with 100, 50 and 25 neurons; activation fn. = relu; Solver = adam; alpha = 0.001 for models 2-var, 3-var, 9-var; alpha = 0.001 for models 1-var and 7-var; adaptive learning rate
Ordinary Least Squares Regression (OLR)	No control parameters required
Random Forest (RF)	Number of estimators = 750; maximum depth = 150; splitting criteria = gini
Stochastic Gradient Descent Classifier (SGD)	Learning rate/regularization α = 0.0001; maximum iterations = 10000
Support Vector Classifier (SVC)	Kernel = rbf; C = 750; gamma = 50 for models 1-var, 2-var, 3-var gamma = 75 for 7-var model; gamma = 25 for 9-var model
XG Boost (XGB)	Number of estimators = 2000; Maximum depth = 7; eta = 0.01; Subsample = 0.7; Columns sampled per tree = 0.5

3.2. Five-fold cross validation analysis

An initial step in the facies classification analysis was to conduct K-fold cross validation for each of the nine models using different training: testing subset splits to establish the splits that generate the most reliable/

reproducible results. Trial and error tests were conducted with different subset splits and the K-fold method. These revealed that the 80%:20% (training: testing) provided the best results for each of the nine models. Hence, the results of the 5-fold cross validation are shown in Table 4.

The 5-fold cross-validation procedure randomly divides the dataset (8911 data records) into five compartments. Four of those compartments are assigned to the training subset (7168 data records) and one to the testing subset (1743 data records). The assignments are repeated five times so that each of the compartments serves as the testing subset in one case. This means that there are five cases evaluated for each run ensuring that each data record is treated as both a training record and a testing record considering all five cases. In this study the 5-fold procedure has been repeated three times, generating results for 15 cases in all. This comprehensive sampling of the dataset provides meaningful statistical assessments (mean and standard deviation of all 15 cases) of the facies class prediction performance of each model evaluated. These 5-fold cross-validation statistics can be usefully compared (Table 4).

Table 4 displays the mean MAE values and MAE's standard deviations (σ^2 MAE) for the 15 cases generated by each ML model. The MAE values are expressed on a scale that relates to facies class (1–5). It is apparent that the are σ^2 MAE low in all cases, and more than an order of magnitude lower than the mean MAE values. This demonstrates that all the models can be relied upon to generate reproducible results with any randomly selected 80:20 split of the dataset into training and testing subsets.

For the 3-var (base case) and 2-var configurations, the SVC model generates the lowest 5-fold mean MAE values, but a higher σ^2 MAE values than some of the other models. For instance, the KNN and RF models generate mean MAE values slightly higher than the SVC model but lower σ^2 MAE values for those variable configurations. Also the XGB model is not among the best performing models for the var-3 and var-2 configurations. However, for the 9-var and 7-var configurations, the XGB model

Table 3

Lithofacies classification performance for the WK-1 Triassic well log section evaluated with nine models for the base case 3-variable (3-var) model applying the optimal 89%:20% split of data records between the training and testing subsets (the testing subsets comprise 1783 of the 8911 data records available).

WK-1 Triassic Section Lithofacies Predictions Using only Recorded Well Log data for GR, DT and PB (3-Variable Model)

Model	Training Subset (80%)				Testing Subset (20%)				Applied to All 8911 Data Records				
	R2	RMSE	MAE	Σ Error	R2	RMSE	MAE	Σ Error	R2	RMSE	MAE	Σ Error	Ex Time
ADA	1.00000	0.00000	0.00000	0	0.57820	0.76153	0.22771	179	0.90795	0.34064	0.04556	179	5.5
DT	1.00000	0.00000	0.00000	0	0.57868	0.76005	0.22434	175	0.90825	0.33998	0.04489	175	5.5
KNN	1.00000	0.00000	0.00000	0	0.66229	0.66522	0.17667	143	0.92912	0.29756	0.03535	143	5.6
MLP	0.54310	0.81124	0.25491	793	0.53628	0.80101	0.25238	202	0.54175	0.80920	0.25440	995	208.2
OLR	0.26040	0.95610	0.69122	2933	0.26708	0.94032	0.67487	724	0.26174	0.95296	0.68795	3657	8.2
RF	1.00000	0.00000	0.00000	0	0.67263	0.65202	0.16938	137	0.93179	0.29166	0.03389	137	75.4
SGD	0.09758	1.19588	0.57295	1957	0.13338	1.12908	0.52328	458	0.10135	1.18765	0.56705	2433	3.4
SVR	0.85652	0.42772	0.07043	222	0.67006	0.65758	0.17106	135	0.81827	0.48255	0.09056	357	29.3
XGB	0.77491	0.54175	0.11448	371	0.46875	0.85650	0.29501	246	0.71032	0.61770	0.15060	617	320.3

Notes: (1)RMSE and MAE are expressed in terms of the facies class range of 1–5; (2)Execution time (Ex Time) is expressed in seconds. It includes the time required for 5-fold cross validation; (3)Error# refers to the number of data records with their facies class misclassified.

Table 4

Five-fold cross validation assessments involving 15 cases of random splits of the dataset between training and testing subsets. Standard deviation (σ^2) is used to indicate the degree of variation among the 15 cases evaluated for each of the nine models evaluated.

5-Fold Cross-Validation Results for Regression and ML Models Applied to 8911 Data Records (15 Cases Run for Each Model)

Model	9 Input Variables		7 Input Variables		3 Input Variables		2 Input Variables		1 Input Variables	
	MAE	σ^2 MAE								
ADA	0.0979	0.0128	0.1864	0.0214	0.2535	0.0154	0.4119	0.0190	0.7928	0.0243
DT	0.0954	0.0127	0.1852	0.0189	0.2537	0.0157	0.4118	0.0213	0.7929	0.0244
KNN	0.0205	0.0060	0.1308	0.0104	0.1973	0.0105	0.3578	0.0151	0.6813	0.0167
MLP	0.1584	0.0142	0.4328	0.0146	0.2620	0.0197	0.3456	0.0206	0.6289	0.0168
OLR	0.6900	0.0085	0.9149	0.0101	0.6880	0.0090	0.7094	0.0087	0.9178	0.0110
RF	0.0307	0.0092	0.0643	0.0109	0.1980	0.0132	0.3498	0.0152	0.7917	0.0243
SGD	0.5575	0.0235	0.6720	0.0150	0.5607	0.0231	0.6234	0.0193	0.6627	0.0205
SVC	0.0501	0.0092	0.2258	0.0213	0.1941	0.0173	0.3309	0.0178	0.6224	0.0154
XGB	0.0245	0.0054	0.0499	0.0068	0.3300	0.0219	0.3894	0.0200	0.6317	0.0162

generates the lowest σ^2 MAE values, together with the lowest mean MAE value for the 7-var configuration, and the second lowest mean MAE value for the 9-var configuration. The KNN, RF and XGB models outperform the other models for the 9-var and 7-var configurations, indicating that the SVC model performs less well with the 9-var and 7-var configurations involving the GR attributes. Not surprisingly, none of the models performs well in lithofacies classification for the var-1 (GR only) configuration in terms of mean MAE. However, the SVC, MLP and XGB models perform better than the other models. The ordinary least squares regression (OLR) model shows the poorest prediction performance for all variable configurations considered. The gradient descent regression models (SGD) performs less well than all the ML models for the 9-var, 7-var, 3-var and 2-var models, although it outperforms several of the ML models for the 1-var case.

In addition to confirming reproducibility of the model evaluated, the key conclusion from the 5-fold cross-validation analysis is that for almost all models their lithofacies prediction performance can be ranked as follows:

(Best) 9-var > 7-var > 3-var > 2-var » 1-var (Worst)

There are two exceptions to this general trend: the SVC and MLP models provide poorer prediction performance for the 7-var configuration than they do for the 3-var configuration. Nevertheless, this general ranking trend observed, based on Table 4 results, confirms that the addition of the calculated GR attributes to the variable configurations is substantially improving their lithofacies prediction performance for the WK-1 Triassic section evaluated.

3.3. Detailed lithofacies prediction results for the testing subsets

A clearer understanding of the relative lithofacies prediction performances of the different models with the different variable configurations can be obtained by considering MAE together with the other prediction performance metric values for a randomly selected testing subset. These results are shown for the four best performing models (KNN, RF, SVC and

Table 5

Lithofacies class prediction performance of the four better performing models applied to the randomly selected 1783 data records (20% of dataset) of the testing subsets for the five different variable configurations evaluated. The results for the other five models with poorer prediction performance are included as Appendix B.

Better Performing ML Model Results for the Testing Subsets					
KNN	9- Var	7-Var	3-Var	2-Var	1-Var
Test R2	0.9741	0.7637	0.6623	0.4527	0.0680
Test RMSE	0.1772	0.5513	0.6652	0.8772	1.2718
Test MAE	0.0146	0.1256	0.1767	0.3186	0.6798
Total Error Number	14	110	143	280	613
Execution Time (seconds)	7.1	6.6	5.6	5.7	5.3
RF					
Test R2	0.9604	0.9078	0.6726	0.4531	0.0462
Test RMSE	0.2196	0.3374	0.6520	0.8791	1.3815
Test MAE	0.0224	0.0499	0.1694	0.3163	0.7835
Total Error Number	23	46	137	274	702
Execution Time (seconds)	113.6	94.3	75.4	61.3	92.3
SVC					
Test R2	0.9248	0.6845	0.6701	0.4603	0.1047
Test RMSE	0.3033	0.6416	0.6576	0.8601	1.1898
Test MAE	0.0393	0.1884	0.1711	0.3034	0.6046
Total Error Number	37	197	135	262	580
Execution Time (seconds)	20.2	52.7	29.3	83.1	192.7
XGB					
Test R2	0.9609	0.9351	0.4687	0.4098	0.0962
Test RMSE	0.2183	0.2822	0.8565	0.9160	1.2080
Test MAE	0.0219	0.0404	0.2950	0.3567	0.6259
Total Error Number	22	45	246	326	605
Execution Time (seconds)	486.8	466.5	320.3	321.0	62.5

XGB) in Table 5.

Those results for the other poorer performing models are shown in Appendix B Table B1. Considering the four performance metrics (R^2 , RMSE MAE and \sum Error) together provides a clearer indication of how each model is performing with respect to the modelled dataset. The relationships between these performance measures are shown in Figs. 7–9.

It is apparent from Table 4 that KNN generates the best lithofacies prediction performance for the 9-var configuration in terms of all error metrics, resulting in only 14 prediction errors among the 1743 data records of the testing subset. For the 7-var configuration XGB outperforms the other models in terms of all error metrics. For the 3-var model SVC outperforms the other models with the lowest \sum Error. However, the RF model generates slightly lower RMSE and MAE values, and a slightly higher R^2 value than the SVC model. The SVC model outperforms the other models for the var-2 and var-1 configurations. The relatively poor performance of the SVC model for the 7-var model suggest that the PB log (the one displaying the highest R and p values with facies class) need to be involved in the variable configuration for the SVC model to perform well.

Fig. 7 reveals that there is a well-defined trend of increasing \sum Error with MAE among the different models evaluated. Excluding configuration var-1, the OLR and SGD model deliver poorer prediction performance than the ML models for all other variable configurations.

Fig. 8 displays, on an expanded scale, only the high performing end of the trend shown in Fig. 7. That figure confirms that the ML models (excluding the MLP and SVC for var-7 configuration) all generate improved facies class prediction performance for var-9 and var-7 configurations compared to the var-3 configuration. Fig. 9 reveals the same result for R^2 versus RMSE with the ML models following a well-defined trend.

The results presented in Table 4 and Figs. 7–9 therefore confirm the findings of the 5-fold cross validation analysis (Table 3). This finding is very encouraging from the perspective of the effectiveness of calculated GR attributes in improving lithofacies class prediction applying ML models to a limited set of well logs. Of particular significance are the var-7 model results because that configuration only uses the recorded GR log data plus the calculated GR attributes. Yet the var-7 configuration provides significantly better facies class prediction results than the 3-var model involving the three recorded logs GR, DT and PB. A key reason this result is that in many wells only a limited range of well logs are available. Indeed, sometimes only the GR log is recorded over certain intervals. The results indicate that reliable lithofacies classification can be achieved using only the recorded GR log and the six GR-attribute variables for the WK-1 Triassic section.

4. Discussion

4.1. Confusion analysis of the best performing ML models

It is useful to understand in more detail the distributions of facies prediction errors generated by the high performing models. Annotated confusion matrices provide useful insight in this regard by considering the classification error metrics defined in Appendix 1. The value that such displays provide is illustrated by Fig. 10 showing confusion matrices for the high performing RF model applied to all 8911 data records of the 9-var, 7-var and 3-var models. For the 9-var configuration the RF model generates only 23 classification errors, and 15 of those errors are associated with misclassification of facies class 2 (Fig. 10A). That model generates no misclassifications for facies class 1 (limestone) and only 1 error for facies class 4 (halite/evaporites). However, accuracy, precision, recall and F1 score are high for all classes.

For the 7-var configuration, the RF model generates just 47 misclassifications (Figs. 10B), 34 of which are misclassifications of facies class 2 (confused mainly with classes 3 and 5). Eleven of the remaining misclassification errors are associated with data records from other classes being confused as facies class 2. However, accuracy, precision,

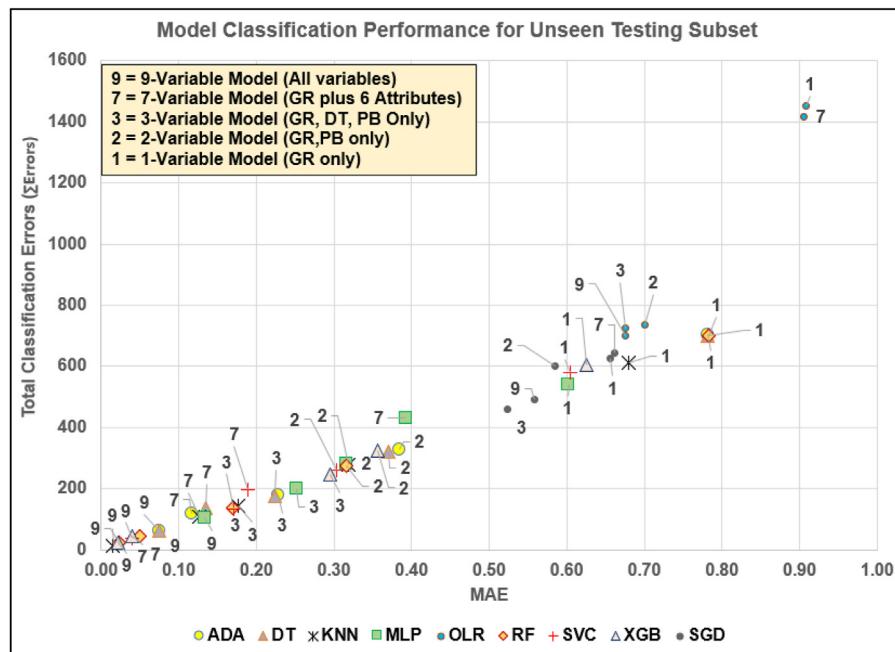


Fig. 7. Σ Error versus MAE for the testing subsets of nine prediction models applied to the five input variable configurations evaluated, labelled 9, 7, 3, 2, and 1.

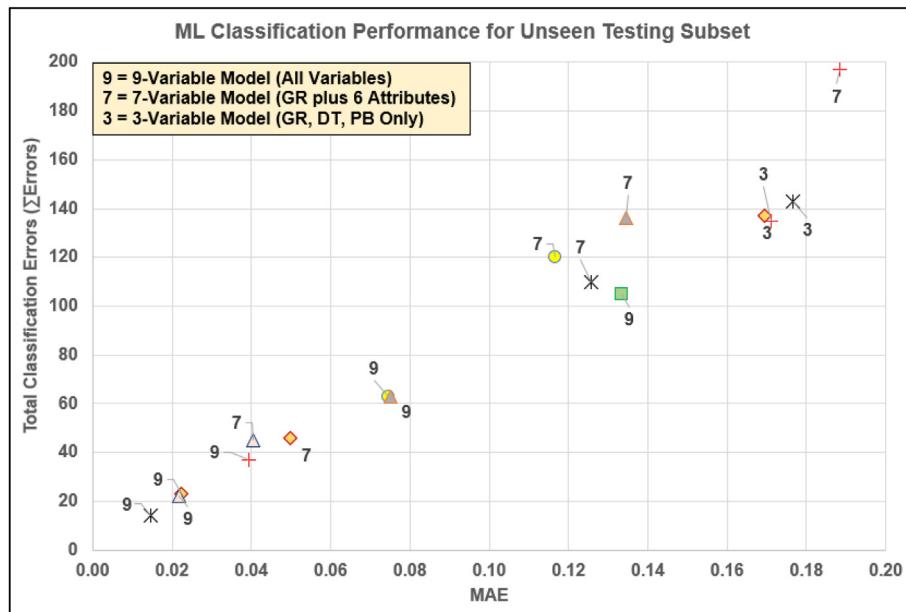


Fig. 8. Σ Error versus MAE for the testing subsets of the best performing classification models (Var-9, Var-7 and Var-3). Symbols are the same as those used in Fig. 6.

recall and F1 score remain above 0.99 for all classes, indicating excellent prediction performance.

For the 3-var configuration, the RF model generates 137 misclassifications (Figs. 10C), 67 of which are misclassifications of facies class 2 (confused mainly with classes 3 and 5). Forty-one of the remaining misclassification errors are associated with data records from other classes being confused as facies class 2. The main deterioration in prediction performance from the 9-var and 7-var configurations is that there are many more confusions between facies class 2 (dolomitic mudstone) and facies class 5 (shaly sandstone). The precision, recall and F1 score fall

substantially below 0.99 for class 5, whereas accuracy, precision and F1 score fall slightly below 0.99 for class 2. The classification performance of the RF model for the 3-var configuration remains very good (98.5% successful predictions) but is clearly inferior to the 9-var and 7-var models due primarily to increased confusion between facies classes 2 and 5. It is also informative to plot the interpreted versus predicted facies class for all data points. Fig. 11 presents such displays for the most accurate ML models (KNN; XGB; RF), those generating the least misclassification, for configurations 9-var, 7-var and 3-var, respectively.

Fig. 11A displays the results of each data record for the KNN model

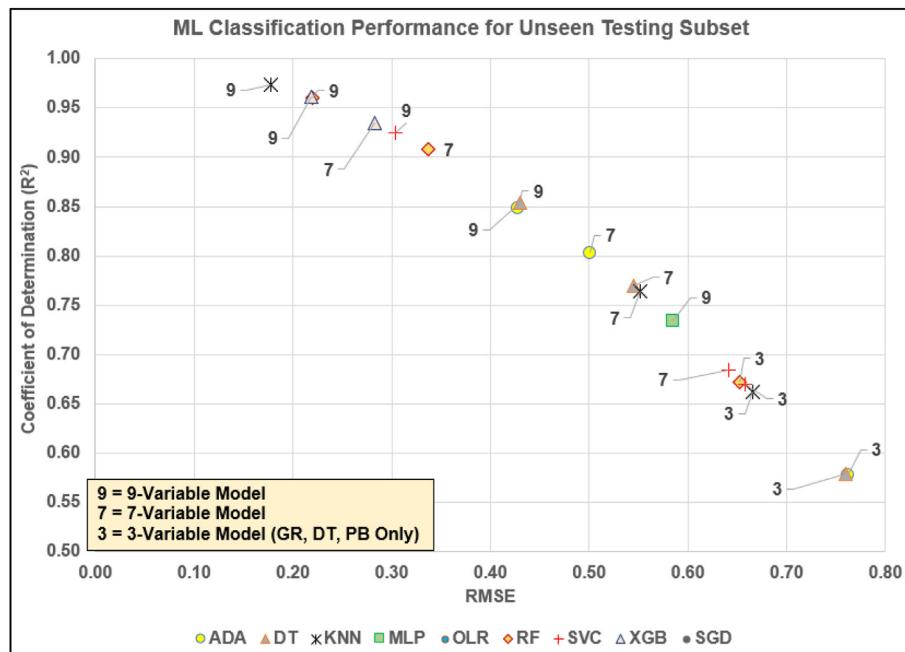


Fig. 9. R^2 versus RMSE for the testing subsets for best performing classification models (Var-9, Var-7 and Var-3).

applied to the 9-var model, highlighting the impressive prediction performance achieved (just 14 errors among 8911 data records). Fig. 11B reveals, from the results of each data record for the XGB model applied to the 7-var model, more misclassifications of other classes as class 2, particularly for class 5 and class 3. Fig. 11C, from the results of each data record for the RF model applied to the 3-var model, highlights and confirms the misclassifications already identified by the confusion matrix shown in Fig. 10C.

Figs. 10 and 11 explain why the inclusion of the GR attributes improves facies class prediction performance for the WK-1 Triassic section. They do so specifically by distinguishing facies classes 2 and 5 with greater precision.

Fig. 12 provides the detailed analysis of the misclassifications for the best-performing (KNN) model for the 9-var configuration that involves just fourteen misclassified data records. Eight of the fourteen misclassifications are data records from other lithofacies classes wrongly classified as lithofacies class 2. Six of the misclassifications occur within 1 m of an actual lithofacies boundary, which are in the lithofacies transition zones that are the hardest for ML algorithms to classify correctly.

4.2. Implications of results and additional studies required

Calculating GR attributes can clearly be beneficial in improving multiple lithology facies classifications from wells with limited well log variables recorded. They add additional perspectives (features) to the GR data that cannot be easily discerned visually. Those additional features may not correlate highly with the facies classes (Fig. 5) but can be effectively used by several ML models to improve their classifications. Significantly, they can do this for the multiple lithology section evaluated when used only in combination with the recorded GR log (the 7-var model). This performance has the potential to be exploited for lithofacies classification in wellbores where GR is the only recorded well log available, which is the case in many top hole sections of many field development wells drilled. This author is not suggesting that GR and its attributes be used in isolation for lithofacies classification as a matter of preference. For detailed facies analysis a suite of diverse data inputs is

always preferable, including mineralogy, various core analysis, fossil and fauna characterization, as well as multiple recorded well logs. Despite, such inputs being desirable, there are many wellbores for which such data is not available due to both cost and technical constraints. In such instances, the GR log plus attributes could offer the best alternative available to provide reliable basic lithofacies characterization with the assistance of an ensemble of ML models.

The KNN model provides the best prediction results with the fewest misclassifications for the 9-var configuration. However, KNN is outperformed by the RF and XGB models for the 7-var configuration (including GR and the GR attributes only). Unlike the other ML models evaluated, the KNN model uses data matching rather than establishing complex relationships between the input variables and dependent variable when formulating its lithofacies predictions. For the dataset evaluated KNN's performance improves progressively as more input data variables become available (Table 5). KNN's correlation-free, data-matching technique seems better able to exploit the additional variations provided in the 9-var configuration, by adding the PB and DT data distributions to the 7-var configuration, than the RF and XGB models.

An observed outcome worthy of further consideration is the relatively poor performance of the MLP and SVC models for the 7-var configuration. The 7-var configuration is the only one not to involve the recorded PB log, the one recorded log that correlates reasonable with facies class; $R = -0.49$ and $p = -0.54$ (Fig. 5). It seems that in the absence of the PB data the MLP and SVC models' prediction performances decline substantially. It is the author's experience with other datasets in which the input variables are all poorly correlated with the dependent variable that the neural-network-based and regression-based models tend to perform much worse in their predictions they generate than tree-based models (such as ADA, DT, RF and XGB) and regression-free data-matching models (e.g. KNN). This may explain the relatively poor performance of the MLP and SVC models with the 9-var and 7-var configurations of the WK-1 dataset. However, more research with other logged lithofacies sections is required to confirm this possibility.

On the other hand, the XGB model shows the reverse prediction performance trend to the MLP and SVC models. XGB performs extremely

(A) Random Forest 9-Variable Model (8911 Data records)						
8911	Formula	249	4745	1593	1235	1089
Total	Class	1	2	3	4	5
251	1	249	2	0	0	0
4735	2	0	4730	1	1	3
1600	3	0	9	1590	0	1
1235	4	0	1	0	1234	0
1090	5	0	3	2	0	1085
Precision: 100.00% 99.68% 99.81% 99.92% 99.63%						
23	Errors	0	15	3	1	4
99.74% Successful Predictions						
Class Prediction Performance:						
Accuracy		0.9998	0.9978	0.9985	0.9998	0.9990
Precision		1.0000	0.9968	0.9981	0.9992	0.9963
Recall		0.9920	0.9989	0.9938	0.9992	0.9954
Balanced F1 Score		0.9960	0.9979	0.9959	0.9992	0.9959
(B) Random Forest 7-Variable Model (8911 Data records)						
8911	Formula	251	4758	1584	1233	1085
Total	Class	1	2	3	4	5
251	1	249	2	0	0	0
4735	2	2	4724	0	4	5
1600	3	0	16	1583	0	1
1235	4	0	6	0	1229	0
1090	5	0	10	1	0	1079
Precision: 99.20% 99.29% 99.94% 99.68% 99.45%						
47	Errors	2	34	1	4	6
99.47% Successful Predictions						
Class Prediction Performance:						
Accuracy		0.9995	0.9949	0.9980	0.9989	0.9981
Precision		0.9920	0.9929	0.9994	0.9968	0.9945
Recall		0.9920	0.9977	0.9894	0.9951	0.9899
Balanced F1 Score		0.9920	0.9953	0.9943	0.9959	0.9922
(C) Random Forest 3-Variable Model (8911 Data records)						
8911	Formula	247	4761	1601	1233	1069
Total	Class	1	2	3	4	5
251	1	247	3	0	0	1
4735	2	0	4694	8	6	27
1600	3	0	23	1571	0	6
1235	4	0	8	0	1227	0
1090	5	0	33	22	0	1035
Precision: 100.00% 98.59% 98.13% 99.51% 96.82%						
137	Errors	0	67	30	6	34
98.46% Successful Predictions						
Class Prediction Performance:						
Accuracy		0.9995	0.9878	0.9933	0.9984	0.9900
Precision		1.0000	0.9859	0.9813	0.9951	0.9682
Recall		0.9841	0.9913	0.9819	0.9935	0.9495
Balanced F1 Score		0.9920	0.9886	0.9816	0.9943	0.9588

Fig. 10. Annotated confusion matrices displaying misclassification error analysis for the 9-var, 7-var and 3-var models evaluated with the random forest model for all data records.

well with the 9-var and 7-var configurations (including the GR attributes) but relatively less well for the 3-var and 2-var than the other high performing models, such as RF and KNN (Table 4). It is not clear why this should be the case but it implies that the XGB model performs better when it has a larger number of features available to exploit. Once again, more research with other logged lithofacies sections is required to confirm this possibility.

The fact that different ML models achieve the best classification

performance with different configurations (i.e., KNN with 9-var, XGB with 7-var, RF/SVC with 3-var, SVC/RF/MLP with 2-var) suggests that it is prudent to evaluate an ensemble of ML methods to establish the best performers with each specific dataset and variable configurations. It appears that one specific ML model cannot be relied upon in isolation to always provide the best prediction results. As most of these models can be executed in a few seconds on laptop computers for medium sized datasets, evaluating an ensemble of ML models is typically not too onerous a task in terms of computational effort (Table 2).

The ML models evaluated in this study all, to a degree, lack transparency when it comes to revealing exactly how the facies class predictions are generated for each data record. For many applications this is not a major disadvantage. However, for outlier analysis and detailed data mining more transparent algorithms can provide advantages by revealing specific relationships between certain groups of data records. The transparent open box (TOB) ML algorithm (Wood, 2018) has demonstrated its ability to both predict and data mine lithofacies datasets (Wood, 2019). Future studies are planned to apply TOB lithofacies models to provide more details of how the GR attributes are used in specific lithofacies predictions.

Further work is required and planned to evaluate how effective GR and other well log derivative and volatility attributes are at improving lithofacies classification performance across logged sections in multiple wellbores. This would involve training and validation prediction models with and without attributes in some wells and testing them with unseen data records in offset wells drilled through the same stratigraphic sequence. That work is underway based on the encouraging results achieved in this study.

5. Conclusions

Gamma ray (GR) log attributes offer useful features for improving lithofacies predictions from well logs in boreholes penetrating multiple lithologies but with limited well-log variables recorded. Many wellbores fall into this category, such as development wells for which the sections drilled above the reservoir target typically have minimal well logs recorded (e.g., GR logs only in many cases). Derivative and volatility GR attributes are of particular value because they extract shape characteristics from GR curves that are not captured by the recorded log values. The first and second derivatives plus a moving average of the first derivative calculated over a relevant depth interval complement the recorded GR data. Similarly, the instantaneous volatility, the standard deviation of that volatility plus a moving average volatility over a relevant depth interval also provide distinctive complementary features related to the changes in GR curve shape with depth.

A logged Triassic section drilled in the Wessex Basin (England; well WK-1) penetrating diverse lithologies is used to demonstrate how useful GR attributes relating to its derivatives and volatility components are for improving lithofacies predictions. Using only three recorded well logs (GR, sonic travel time (DT) and bulk density (PB)), regression and machine learning (ML) models struggle to distinguish reliably between some of the five lithofacies identified in the studied wellbore section. Nine models (two linear regression; seven ML) applied to evaluate five distinct log feature configurations (1-var uses the GR curve alone; 2-var uses GR and PB recorded well logs only; 3-var uses GR, DT and PB recorded data only; 7-var uses GR plus six GR attributes; 9-var uses GR, DT, PB and six GR attributes) reveal the substantial benefits to lithofacies prediction of introducing the GR attributes.

The 9-var and 7-var models substantially improve upon the prediction models using only the recorded log data. The best performing 9-var model (K-nearest neighbor (KNN)) generates only 14 facies misclassification from 8911 data records. This compares with 135 misclassification for the best 3-var model (Support Vector Classification (SVC)). Even more

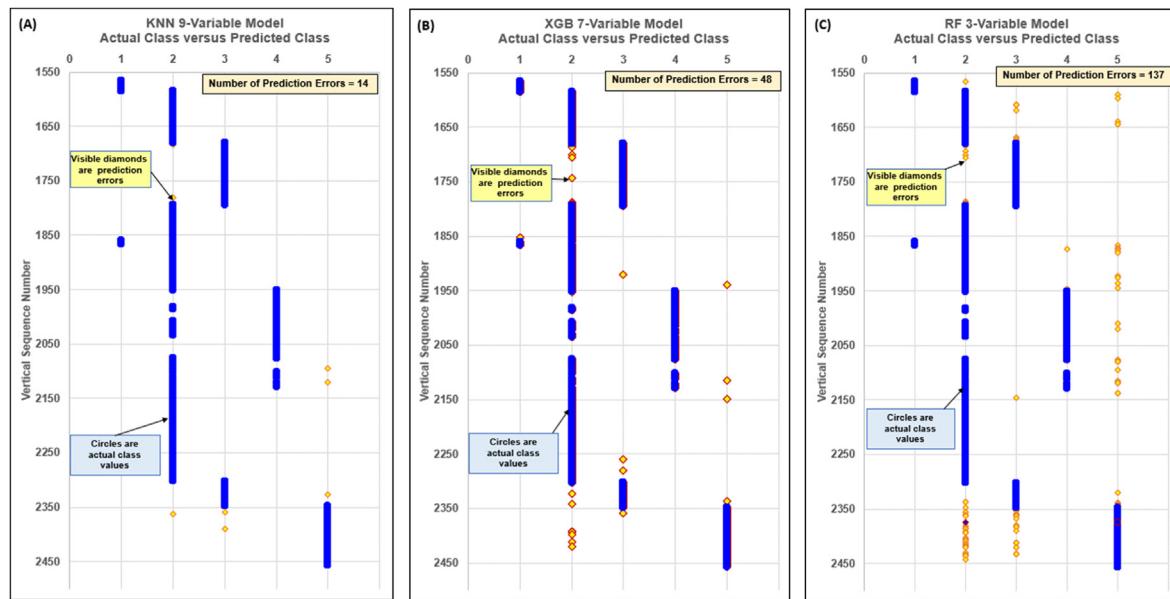


Fig. 11. Observed (blue dots) versus ML-predicted (yellow diamonds) facies lithofacies classes for all 8911 data records of WK-1 Triassic section for the best performing models: (A) KNN for the 9-variable model; (B) XGB for the 7-variable model; and, (C) RF for the 3-variable model. At depths where the observed and prediction lithofacies are in agreement only the observed lithofacies is displayed as a blue dot. At depths where the observed and prediction lithofacies are not in agreement both observed (blue dot) and predicted lithofacies (yellow diamond) are displayed. Figs. 10(C) and 11(C) provide complementary information regarding the performance of the RF 3-variable model.

impressive, as it only involves GR and GR attributes, the best 7-var model (Extreme Gradient Boosting (XGB)) generates only 45 misclassifications. The reproducibility of the ML models is verified by 5-fold cross-validation analysis, which also identifies KNN, Random Forest (RF) and XGB as the best performing ML model with the 9-var and 7-var feature configurations. Annotated confusion matrices reveal that none of the ML models can reliably distinguish facies classes 2 (dolomitic mudstones/siltstones) and 5 (shaly sandstones/siltstones) for the 3-var configuration. Nevertheless, KNN/RF/XGB models manage to do this effectively for the 9-var and 7-var models.

A somewhat surprising outcome of the ML model analysis is the relatively poor performance of the multi-layer perceptron (MLP) and SVC models with the 7-var configuration. That configuration does not include PB data, which is considered significant, as PB is moderately well correlated with facies class (Spearman's correlation coefficient ~ 0.54). All the other features are poorly correlated with facies class, especially the GR attributes. This implies that the MLP and SVC models are less able to deal with more poorly correlated feature selections, although that implication requires further analysis with additional datasets. A less surprising outcome of the analysis is that standard least squares and gradient-descent muti-linear regression models perform substantially worse than ML models for facies classification with all configurations, except the 1-var model. The low correlations between the features,

especially the GR attributes, and facies class explains that outcome. For the logged section studied, the ML models that do not rely on correlations, KNN/RF/XGB, clearly provide superior lithofacies prediction performance. However, one of those models does not consistently outperform the other taking into account the 9-var, 7-var and 3-var configurations. That observation suggests that it is better to apply an ensemble of those three models to ensure optimum lithofacies prediction from a range of feature configurations.

Declaration of interests

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

No funding was received for this study.

Conflicts of interest

The author has no conflicts of interest associated with this study.

Appendix A. Definitions of statistical measures of prediction performance applied

The prediction performance assessment metrics used in this study are defined in Figure A1.

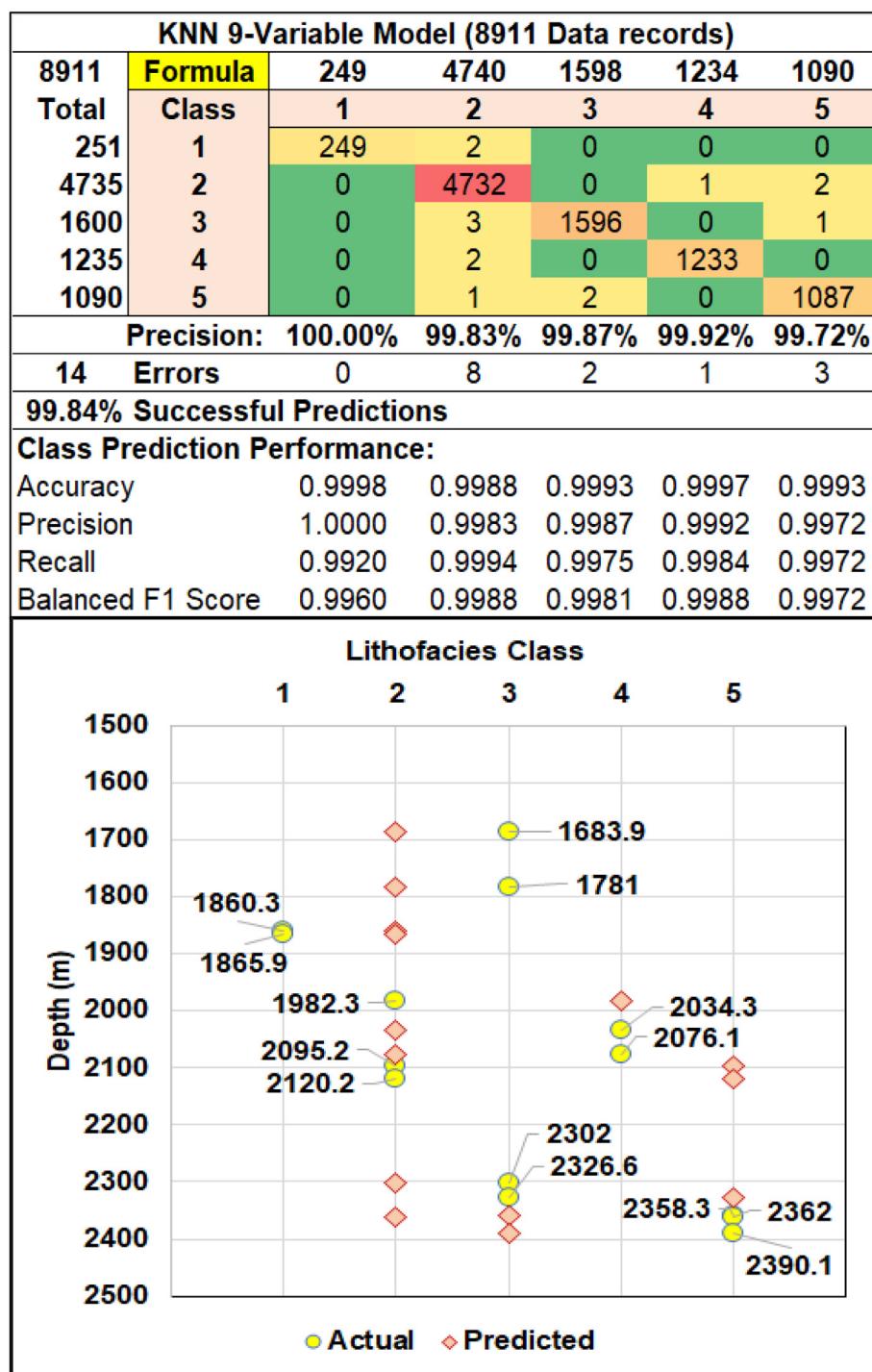


Fig. 12. Misclassification analysis for the best-performing KNN model (KNN) applied to the 9-var dataset configuration. The labels in the lower graphic represent the depth values of the fourteen misclassified data records.

Statistical Measures of Prediction Performance Assessed

Mean Squared Error (MSE): $MSE = \frac{1}{n} \sum_{i=1}^n ((X_i) - (Y_i))^2$

Root Mean Squared Error (RMSE): $RMSE = \sqrt{MSE}$

Mean Absolute Error (MAE): $MAE = \frac{1}{n} \sum_{i=1}^n |X_i - Y_i|$

Correlation Coefficient (R): $R = \frac{\sum_{i=1}^n (Xi - Xmean)(Yi - Ymean)}{\sqrt{\sum_{i=1}^n (Xi - Xmean)^2 \sum_{i=1}^n (Yi - Ymean)^2}}$

Coefficient of Determination (R^2): R is expressed on a scale of -1 to +1
 R^2 is expressed on a scale of 0 to 1

Where:

X_i is the measured value and Y_i is the predicted value for data record i

n is the number of data records in the set or subset being evaluated

D_i is $(Xi - Yi)$ for data record i

$$Dmean = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i)$$

Measures specifically focused on classification errors

Absolute number of prediction errors recorded

Accuracy (A): $A = (TP+TN)/(TP+TN+FP+FN)$

Precision (P): $P = TP/(TP+FP)$

Recall (R): $R = TP/(TP+FN)$

Balanced F-score (F1): $F1 = (2 \times (R \times P))/(R+P)$

Where:

TP = true number of positive predictions, TN = true number of negative predictions,

FP = number of false positive predictions and FN = number of false negative predictions.

These classification prediction error assessments are most usefully displayed and interpreted in conjunction with an annotated confusion matrix.

Fig. A1. Statistical measures of prediction performance appropriate for both regression and classification analysis.

Appendix B. Lithofacies Class Prediction Performance for Additional Models

Lithofacies class prediction performance is reported here for the five models with poorer results overall (Table B1) compared to the KNN, RF, SVC and XGB. Note that the ADA and DT models only slightly underperform the models mentioned. In fact, the ADA and DT models provide better results for the Var-7 configuration than the SVC model, and better results for the Var-3 configuration than the XGB model. The OLR and SGD linear regression models are the worst performing models overall. The MLP model does not perform as well as other ML models evaluated for the Var-9 and Var-7 configurations, but its performance improves relative to other ML models for Var-3 and Var-2 configurations. Indeed, for the Var-1 configuration the MLP model provides the best prediction performance.

Table B1

Lithofacies class prediction performance of the five models with poorer results overall than the KNN, RF, SVC and XGB models. These results are for the randomly selected 1783 data records (20% of dataset) of the testing subsets for the five different variable configurations evaluated. For higher performing model results see [Table 5](#).

Poorer Performing ML Model Results for the Testing Subsets

ADA	9- Var	7-Var	3-Var	2-Var	1-Var
Test R2	0.8480	0.8031	0.5782	0.3663	0.0467
Test RMSE	0.4283	0.5013	0.7615	0.9762	1.3801
Test MAE	0.0746	0.1167	0.2277	0.3842	0.7818
Total Error Number	63	120	179	328	704
Execution Time (seconds)	5.7	5.6	5.5	5.8	199.9
DT					
Test R2	0.8541	0.7694	0.5787	0.3931	0.0467
Test RMSE	0.4302	0.5452	0.7601	0.9520	1.3801
Test MAE	0.0752	0.1346	0.2243	0.3702	0.7818
Total Error Number	63	136	175	321	701
Execution Time (seconds)	5.6	5.7	5.5	5.5	6.1
OLR					
Test R2	0.2730	0.0523	0.2671	0.2535	0.0427
Test RMSE	0.9366	1.0694	0.9403	0.9490	1.0747
Test MAE	0.6760	0.9052	0.6749	0.7003	0.9080
Total Error Number	700	1418	724	736	1453
Execution Time (seconds)	4.9	5.3	5.2	5.0	4.7
MLP					
Test R2	0.7340	0.3943	0.5363	0.4521	0.0937
Test RMSE	0.5849	0.9279	0.8010	0.8672	1.1817
Test MAE	0.1335	0.3932	0.2524	0.3158	0.6012
Total Error Number	105	431	202	282	540
Execution Time (seconds)	210.6	287.6	208.2	129.5	15.2
SGD					
Test R2	0.1074	0.0623	0.1334	0.1516	0.0681
Test RMSE	1.1720	1.2469	1.1291	1.1422	1.2390
Test MAE	0.5580	0.6618	0.5233	0.5855	0.6556
Total Error Number	491	644	458	602	627
Execution Time (seconds)	5.6	3.4	3.4	5.7	3.4

References

- Agrawal, R., Malik, A., Samuel, R., Saxena, A., 2022. Real-Time Prediction of litho-facies from drilling data using an artificial neural network: a comparative field data study with optimizing algorithms. *J. Energy Resour. Technol.* 144, 12. <https://doi.org/10.1115/1.4051573>, 043003.
- Al-Mudhafer, W.J., 2017. Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms. *J. Petrol. Explor. Prod. Technol.* 7, 1023–1033. <https://doi.org/10.1007/s13202-017-0360-0>, 2017.
- Bestagini, P., Lipari, V., Tubaro, S., 2017. A machine learning approach to facies classification using well logs. In: Proceedings of the SEG International Exposition and 87th Annual Meeting, vol. 2137. <https://doi.org/10.1190/segam2017-17729805.1>.
- Bottou, L., 1998. *Online Algorithms and Stochastic Approximations. Online Learning and Neural Networks*. Cambridge University Press, ISBN 978-0-521-65263-6.
- Busch, J., Fortney, W., Berry, L.N., 1987. Determination of lithology from well logs by statistical analysis. *SPE Form. Eval.* 2, 412–418. <https://doi.org/10.2118/14301-PA>.
- Cant, D.J., 1992. Subsurface facies analysis. In: Walker, R.G., James, N.P. (Eds.), *Facies Models, Response to Sea Level Changes*. Geol. Assoc., Canada, pp. 27–45.
- Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Krishnapuram, Balaji, Shah, Mohak, Smola, Alexander J., Aggarwal, Charu C., Shen, Dou, Rastogi, Rajeev (Eds.), *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, San Francisco, CA, USA, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>. August 13–17, 2016.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20 (3), 273–297. <https://doi.org/10.1007/BF00994018>.
- Dubois, M.K., Bohling, G.C., Chakrabarti, S., 2007. Comparison of four approaches to a rock facies classification problem. *Comput. Geosci.* 33, 599–617.
- Emery, D., Myers, K.J., 1996. Sequence Stratigraphy. Blackwell Science, Oxford (U.K.), p. 297. <https://doi.org/10.1002/9781444313710>.
- Fadokun, D.O., Oshilike, I.B., Onyekonwu, M.O., 2020. Supervised and Unsupervised Machine Learning Approach in Facies Prediction. *SPE-203726-MS*. Nigeria Annual International Conference and (Virtual) Exhibition. 2020. <https://doi.org/10.2118/203726-MS>.
- Faga, A.T., Oyeneyin, B.M., 2000. Effects of Diagenesis on Neural-Network Grain-Size Prediction. *SPE-60305-MS*. Rocky Mountain Regional/Low-Permeability Reservoirs Symposium and Exhibition, Denver, Colorado, March 2000. <https://doi.org/10.2118/60305-MS>.
- Farzi, R., Bolandi, V., 2016. Estimation of organic facies using ensemble methods in comparison with conventional intelligent approaches: a case study of the South Pars Gas Field, Persian Gulf, Iran. *Model. Earth Syst. Environ.* 2, 105. <https://doi.org/10.1007/s40808-016-0165-z>.
- Feng, R., Luthi, S.M., Gisolf, D., Angerer, E., 2018. Reservoir lithology classification based on seismic inversion results by hidden Markov models. *Mar. Petrol. Geol.* 93, 218–229. <https://doi.org/10.1016/j.marpgeo.2018.03.004>.
- Feng, R., 2020. Lithofacies classification based on a hybrid system of artificial neural networks and hidden Markov models: applying prior geological information. *Geophys. J. Int.* 221 (3), 1484–1498. <https://doi.org/10.1093/gji/ggaa083>.
- Feng, R., 2021. A Bayesian approach in machine learning for lithofacies classification and its uncertainty analysis. *Geosci. Rem. Sens. Lett. IEEE* 18 (1), 18–22. <https://doi.org/10.1109/LGRS.2020.2968356>.
- Fix, E., Hodges Jr., J.L., 1951. *Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties*. USAF School of Aviation Medicine. Technical Report.
- Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* 55, 119–139. <https://doi.org/10.1006/jcss.1997.1504>.
- Gao, D., 2011. Latest developments in seismic texture analysis for subsurface structure, facies, and reservoir characterization: a review. *Geophysics* 76 (2). <https://doi.org/10.1190/1.3553479>. W1–W13.
- Goldberger, A.S., 1964. *Classical Linear Regression. Econometric Theory*, vol. 158. John Wiley & Sons, New York, ISBN 0-471-31101-4.
- Goncalves, C.A., Harvey, P.K., Lovell, M.A., 1995. Application of a multilayer neural network and statistical techniques in formation characterization. In: SPWLA 36th Annual Logging Symposium, Paris, 26–29 Jun, Society of Petrophysicists and Well Log Analysts, p. 12. Houston (U.S.A.).
- Hall, B., 2016. Facies classification using machine learning. *Lead. Edge* 35 (10), 906–909. <https://doi.org/10.1190/le35100906.1>, 2016.
- Hall, M., Hall, B., 2017. Distributed collaborative prediction: results of the machine learning contest. *Lead. Edge* 36 (3), 267–269. <https://doi.org/10.1190/tle36030267.1>.
- Halotel, J., Demyanov, V., Gardiner, A., 2020. Value of geologically derived features in machine learning facies classification. *Math. Geosci.* 52, 5–29. <https://doi.org/10.1007/s11004-019-09838-0>.
- Ho, T.K., 1998. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (8), 832–844. <https://doi.org/10.1109/34.709601>.
- Hosseini, T.M., Watada, J., Aziz, I.A., Hermana, M., 2020. Machine learning in electrofacies classification and subsurface lithology interpretation: a rough set theory approach. *Appl. Sci.* 10, 5940. <https://doi.org/10.3390/app10175940>.
- Hounslow, M.W., Ruffell, A.H., 2006. Triassic: seasonal rivers, dusty deserts and saline lakes. Chapter in: Brenchley, P.J., Rawson, P.F. (Eds.), *The Geology of England and Wales*. Geological Society of London.

- Howard, A.S., Warrington, G., Ambrose, K., Rees, J.G., 2008. A Formational Framework for the Mercia Mudstone Group (Triassic) of England and Wales. British Geological Survey, Research Report, RR/08/04.
- Hurst, A., 1990. Natural gamma-ray spectroscopy in hydrocarbon bearing sandstones from the Norwegian continental shelf. In: Hurst, A., Lovell, M.A., Morton, A.C. (Eds.), *Geological Applications of Wireline Logs*. Geological Society, London, Special Publication, vol. 48, pp. 211–222.
- Ippolito, M., Ferguson, J., Jenson, F., 2021. Improving facies prediction by combining supervised and unsupervised learning methods. *J. Petrol. Sci. Eng.* 200, 108300. <https://doi.org/10.1016/j.petrol.2020.108300>.
- Kaya, M., 2015. Provenance and migration of the petroleum fluids in the Wessex basin, Southern England. *Bull. Turk. Assoc. Petrol. Geol.* 27 (1), 31–56.
- Kesslar, L., Sachs, S.D., 1995. Depositional Setting and Sequence Stratigraphic Implications of the Upper Sinemurian (Lower Jurassic) Sandstone Interval, 1995, vol. 93. North Celtic Sea/St George's, Channel Basins, offshore Ireland. *Geol Soc Lond Spec Publ*, pp. 171–192. <https://doi.org/10.1144/GSL.SP.1995.093.01.13>, 1.
- Kim, Y., Hardisty, R., Torres, E., Marfurt, K.J., 2018. Seismic facies classification using random forest algorithm. In: SEG International Exposition and 88th Annual Meeting, pp. 2161–2165. 0.1190/segam2018-2998553.1.
- Krumbein, W.C., Sloss, L.L., 1951. Stratigraphy and Sedimentation: San Francisco. W.H.Freeman and Co, p. 497. <https://doi.org/10.1002/gj.3350010110>, 1951.
- Mandal, P.P., Rezaee, R., 2019. Facies classification with different machine learning algorithm – an efficient artificial intelligence technique for improved classification. *ASEG Extended Abstracts* 1, 1–6. <https://doi.org/10.1080/22020586.2019.12072918>.
- Merembayev, T., Kurmangaliyev, D., Bekbauov, B., Amanbek, Y.A., 2021. Comparison of machine learning algorithms in predicting lithofacies: case studies from Norway and Kazakhstan. *Energies* 14. <https://doi.org/10.3390/en14071896>, 1896.
- Newell, A.J., Woods, M.A., Graham, R.L., Christodoulou, V., 2021. Derivation of lithofacies from geophysical logs: a review of methods from manual picking to machine learning. In: *British Geological Survey (Nottingham, England) Open Rep. OR/21/006*. 43pages.
- Potratz, G.L., Canchumuni, S.W.A., Castro, J.D.B., Potratz, J., Pacheco, M.A.C., 2021. Automatic Lithofacies Classification with T-SNE and K-Nearest Neighbors Algorithm, vol. 44. Anuário do Instituto de Geociências, 35024. https://doi.org/10.11137/2021_44.35024.
- Puskarczyk, E., 2019. Artificial neural networks as a tool for pattern recognition and electrofacies analysis in Polish palaeozoic shale gas formations. *Acta Geophys.* 67, 1991–2003. <https://doi.org/10.1007/s11600-019-00359-2>.
- Quinlan, J.R., 1986. Induction of decision trees. *Mach. Learn.* 1, 81–106. <https://doi.org/10.1007/BF00116251>.
- Reverdy, X., Argaud, M., Walgenwitz, F., 1983. Mineralogical Analysis Required for Log Interpretation in Complex Lithologies. Paper H, *Transactions of the SPWLA 8th European Symposium*.
- Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Cornell aeronautical laboratory. *Psychol. Rev.* 65 (6), 386–408. <https://doi.org/10.1037/h0042519>.
- Rhys, G.H., Lott, G.K., Calver, M.A., 1982. The Winterborne Kingston Borehole, Dorset, England. Rep CF81/03. Stationery Office Books, p. 196, 13:978-0118841931, London, United Kingdom.
- Rider, M.H., 1986. *Geological Interpretation of Well Logs*. Blackie, ISBN 9780216918467, p. 175. New York (U.S.A.).
- Rider, M.H., 1990. Gamma-ray log shape used as a facies indicator: critical analysis of an oversimplified methodology. *Geol. Soc. Lond. Spec. Publ.* 48, 27–37. <https://doi.org/10.1144/GSL.SP.1990.048.01.04>.
- Rogers, S.J., Fang, J., Karr, C., Stanley, D., 1992. Determination of lithology from well logs using a neural network (I). *AAPG (Am. Assoc. Pet. Geol.) Bull.* 76, 731–739.
- Russell, W.L., 1944. The total gamma ray activity of sedimentary rocks as indicated by Geiger counter determinations. *Geophysics* 9 (2), 180–216. <https://doi.org/10.1190/1.1445076>.
- Sarkar, S., Majundar, C., 2020. A comparative analysis of supervised classification algorithms for lithofacies characterization. In: EAGE Digitalization Conference and Exhibition, Nov 2020. European Association of Geoscientists & Engineers, pp. 1–5. <https://doi.org/10.3997/2214-4609.202032090>.
- Scholle, P.A., Spearing, D., 1982. *Sandstone Depositional Environments*, vol. 31. American Association of Petroleum Geologists Memoir, ISBN 9780891813071, p. 410.
- SciKit Learn, 2022a. Supervised and Unsupervised Machine Learning Models in Python, 2022a. <https://scikit-learn.org/stable/>, 2nd February 2022.
- SciKit Learn, 2021a. GridSearchCV: Exhaustive Search over Specified Parameter Values for an Estimator in Python. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html, 2nd February 2022.
- SciKit Learn, 2021b. Bayesian Optimization of Hyperparameters in Python. <https://scikit-optimize.github.io/stable/modules/generated/skopt.BayesSearchCV.html>, 2nd February 2022.
- SciKit Learn, 2022b. Cross-validation: Evaluating Estimator Performance. https://scikit-learn.org/stable/modules/cross_validation.html, 2nd February 2022.
- Selley, R.C., 1978. Concepts and Methods of Subsurface Facies Analysis. American Association of Petroleum Geologists Education Course Notes Series #9. <https://doi.org/10.1306/CE9397>.
- Shashank, S., Mahapatra, M.P., 2018. Boosting Rock Facies Prediction: Weighted Ensemble of Machine Learning Classifiers. Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 2018, SPE-192930-MS. <https://doi.org/10.2118/192930-MS>.
- Sun, Z., Jiang, B., Li, X., Li, J., Xiao, K., 2020. A data-driven approach for lithology identification based on parameter-optimized ensemble learning. *Energies* 13 (15), 3903. <https://doi.org/10.3390/en13153903>.
- Tran, T.V., Ngo, H.H., Hoang, S.K., Tran, H.N., Lambiase, J.L., 2020. Depositional Facies Prediction Using Artificial Intelligence to Improve Reservoir Characterization in a Mature Field of Nam Con Son Basin, Offshore Vietnam. Offshore Technology Conference Asia, Kuala Lumpur, Malaysia, November 2020 OTC-30086-MS. <https://doi.org/10.4043/30086-MS>.
- Underhill, J.R., Stoneley, R., 1998. Introduction to the development, evolution and petroleum geology of the Wessex Basin. In: Underhill, J.R. (Ed.), *Geol. Soc. Lond. vol. 133. Spec. Publ.*, pp. 1–18.
- Wood, D.A., 2018. Transparent Open-Box learning network provides insight to complex systems and a performance benchmark for more-opaque machine learning algorithms. *Adv. Geo-Energy Res.* 2 (2), 148–162. <https://doi.org/10.26804/ager.2018.02.04>.
- Wood, D.A., 2019. Lithofacies and stratigraphy prediction methodology exploiting an optimized nearest-neighbour algorithm to mine well-log data. *Mar. Petrol. Geol.* 110, 347–367. <https://doi.org/10.1016/j.marpetgeo.2019.07.026>.
- Wood, D.A., 2020. Bakken stratigraphic and type well log learning network for transparent prediction and rigorous data mining. *Nat. Resour. Res.* 29 (2), 1329–1349. <https://doi.org/10.1007/s11053-019-09525-1>.
- Wood, D.A., 2022. Gamma-ray log derivative and volatility attributes assist facies characterization in clastic sedimentary sequences for formulaic and machine learning analysis. *Adv. Geo-Energy Res.* 6 (1), 69–85. <https://doi.org/10.46690/ager.2022.01.06>.
- Wrona, T., Pan, I., Gawthorpe, R.L., Fossen, H., 2018. Seismic facies analysis using machine learning. *Geophysics* 83 (5), O83–O95. <https://doi.org/10.1190/GEO2017-0595.1>.
- Xie, Y., Zhu, C., Lu, Y., Zhu, Z., 2019. Towards optimization of boosting models for formation lithology identification. *Math. Probl Eng.*, 5309852 <https://doi.org/10.1155/2019/5309852>.