

Original article

Gamma-ray log derivative and volatility attributes assist facies characterization in clastic sedimentary sequences for formulaic and machine learning analysis

David A. Wood[✉]*

DWA Energy Limited, Lincoln, United Kingdom

Keywords:

Gamma-ray derivatives
volatility attributes
clastic facies characterization
formulaic facies discrimination
machine learning
feature augmentation

Cited as:

Wood, D. A. Gamma-ray log derivative and volatility attributes assist facies characterization in clastic sedimentary sequences for formulaic and machine learning analysis. *Advances in Geo-Energy Research*, 2022, 6(1): 69-85. <https://doi.org/10.46690/ager.2022.01.06>

Abstract:

Rate of change, second derivative and volatility of gamma-ray (GR) well-log curves provide useful indicators with which to characterize lithofacies in clastic sedimentary sequences. Rolling averages of these variables, as they change with depth, are also able to distinguish certain lithofacies features. These attributes make it possible to accurately distinguish key facies by using only gamma-ray data, both with formulaic calculations and employing machine-learning (ML) algorithms. This is useful in the many wellbores for which only basic logging suites are available. As well as enhancing lithofacies classification more generally using well-log variables, these GR attributes can be used to forecast facies in real time based on logging-while-drilling data. The application is demonstrated with simple formula using synthetic GR logs featuring common clastic lithofacies and their transitions. Seven widely used ML methods are each trained and validated with a synthetic GR curve (1450 data points) displaying six distinct facies. The ability of the ML model to distinguish those facies using seven GR attributes is compared and further tested with an independent GR data set (800 data points). The random forest algorithm outperforms the other ML models in this facies prediction task, achieving a mean absolute error of 0.25 (on a facies class range of 1 to 6) for the independent testing dataset. The results highlight the benefit of this technique in providing reliable facies analysis based only on GR data. Random forest, support vector classification and eXtreme Gradient Boost are the ML models that provide the most reliable facies classification from the GR attributes defined. Annotated confusion matrices assist in revealing the details of facies class prediction accuracy and precision achieved by the ML and models and classification formulas.

1. Introduction

Interpreting lithofacies and inferring depositional environments from the shape of well-log curves has been successfully applied in many sedimentary basins around the world, dating back to the pre-1950's (Krumbein and Sloss, 1951; Selley, 1978). Indeed, interpreting the shape of basic well logs, gamma ray (GR) (Russell, 1944), spontaneous potential, and acoustic travel time, has contributed to the fundamental interpretation of sandstone depositional environments for many decades (Scholle and Spearing, 1982). Well-log shape is particularly useful for characterizing clastic lithofacies (Rider, 1986), and is also a key feature of sequence stratigraphy

(Emery and Myers, 1996). Cant (1992) defined some generic well-log-curve shapes associated with sandstones (Fig. 1). GR-log curves have also been exploited to interpret grain-size changes with depth (Hurst, 1990; Faga and Oyenehin, 2000). This is possible because the radioactive-rich shale contents tend to be greater in the more finely grained clastic sediments (Dypvik and Eriksen, 1983). Additionally, as clay/shale content of clastic sediments decrease the GR curve progressively deflects towards the left on the log trace as their silt and then sand contents increase (Rider, 1990; Kessler and Sachs, 1995). These features make the GR curves highly responsive to facies changes in clastic sequences. However, in complex lithologies

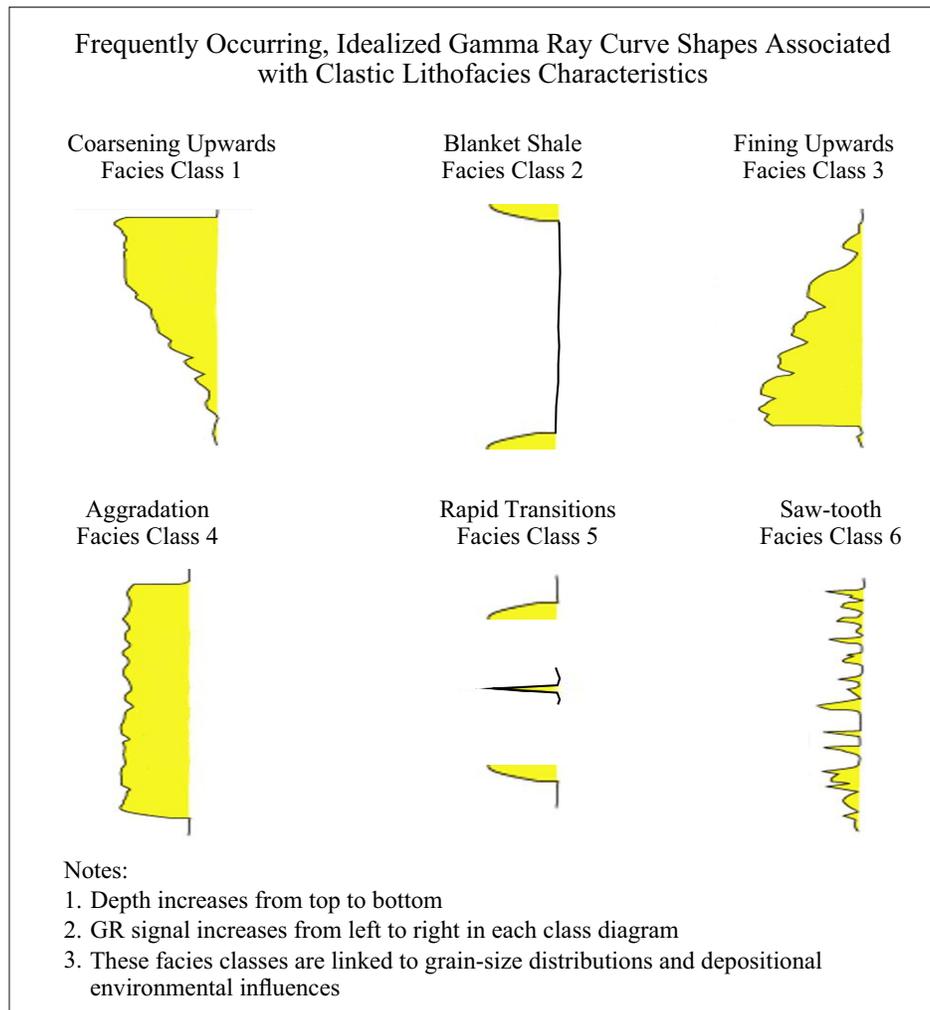


Fig. 1. GR curve shapes versus depth diagrammatically illustrating six commonly encountered lithofacies features in clastic sequences. The class numbers are those attributed to the GR sequences evaluated in this study.

additional mineralogical information, from cores or additional logs, is required for detailed lithofacies interpretation from GR data (Reverdy et al., 1983).

Following earlier computational efforts to predict lithofacies from well log data (Busch et al., 1987), neural networks have been widely applied since the 1990's to assist in lithofacies predictions from well log datasets (Rogers et al., 1992). Using multiple well logs, Dubois et al. (2007) demonstrated that neural networks could outperform k-nearest neighbor, Bayesian and fuzzy classifiers when applied to a large dataset from the Panoma gas field (Kansas, U.S.A.). Bestagini et al. (2017) achieved improved lithofacies prediction from a multi-well-log data set by applying feature augmentation, specifically quadratic expansion and combining information from adjacent formations, and a gradient boosting algorithm. Wood (2019, 2020), also with a suites of multiple well logs, applied an optimized data matching algorithm combined with data upscaling, to successfully distinguish lithofacies in the Hassi R'Mel gas field (Algeria) and the Bakken Province (U.S.A.).

Clustering behavior of GR data, in isolation, was used to

distinguish clean reservoir sands zones with the aid of a multi-layered neural network (Goncalves et al., 1995), representing GR's potential use in machine learning (ML) applications to characterize sandstone lithofacies. In recent years several different types of ML methods have been applied for facies interpretation using well log data incorporating GR curves, but have met with relatively limited success (Hall, 2016). This has led many to conclude that geological information from core analysis and/or biostratigraphy are necessary features to be included for reliable machine learning lithofacies classification (Halotel et al., 2020; Tran et al., 2020). Combining supervised and unsupervised machine learning with clustered GR data has recently become a popular tool for lithofacies analysis (Fadokun et al., 2020; Ippolito et al., 2021).

This study explores, in a novel way, the potential of feature augmentation to GR data for the purpose of automated lithofacies characterization of clastic sequences, using only GR curve data as the primary data source. Specifically, first and second derivatives (i.e., rate of change and rate of rate of change) and the rolling average of the first derivative of the GR data are established. The derivative data is used together with

the volatility attributes of the GR data versus depth as input variables with which to distinguish and classify commonly encountered clastic lithofacies features. The method is applied both in terms of discriminatory formulas and using several standard ML classification methods and the results compared.

2. Method and materials

2.1 Lithofacies characteristics considered

Six lithofacies characteristics are investigated (Fig. 1). These include classic GR-log-curve shapes widely described and studied (e.g., Cant, 1992).

These lithofacies features are assigned class numbers 1 to 6 as follows:

Class 1: A coarsening-upwards sequence, typically resulting in a funnel-shaped GR-log profile. This profile is commonly associated with prograding depositional environments, such as those found in river-mouth fronts, delta fronts and shore faces. The upper parts of such features (low GR values) often have good porosity and constitute prolific oil/gas reservoirs. Such features are therefore important to identify from an exploration perspective. They are typically associated with a relatively abrupt (steep GR curve) upper boundary.

Class 2: A thick, relatively homogeneous shale layer, characterized by high and essentially flat GR values. These are often associated with transgressive depositional environments in which the shallower water environments of the underlying sandstones are replaced for a substantial period by deeper water conditions dominated by persistent shale deposition.

Class 3: A fining-upwards sequence, typically resulting in a bell- or Christmas-tree shaped GR-log profile. This profile is commonly associated with retrograding depositional environments, such as those found in river bars, delta tributaries and some proximal marine environments. The lower portion of such features (low GR values) often have good porosity and constitute prolific oil/gas reservoirs. Such features are typically associated with a relatively abrupt (steep GR curve) lower boundary.

Class 4: A massive, often quite thick, clean sand (low GR values throughout), typically resulting in a cylindrical- or barrel-shaped GR profile. This profile is commonly associated with aggrading depositional environments, such as those found in fluvial channels, aeolian canyons, and major delta distributary channels during persistent progradation. Such massive clean sandstones often have good porosity and can constitute thick and prolific oil/gas reservoirs. They can be associated with relatively abrupt (steep GR curve) or gradational (less inclined GR curves) at their lower and upper boundaries.

Class 5: An abrupt transition zone from one facies to another occurring over a small, but not insignificant, depth interval. This is characterized by a very steep GR curve, sloping upwards or downwards. In manual facies analysis such zones might not be distinguished for separate consideration. However, in automated ML analysis, it is often helpful to treat such zones as separate features. They may form upper or lower boundaries to features described as classes 1 to 4. Additionally, they may be associated with thin sandstone and/or siltstone interbeds within a more massive shaly sequence (either class

4 or class 6).

Class 6: A serrated or irregular, saw-tooth rapidly changing GR profile, typically dominated by high but variable GR values interspersed erratically with thin low-GR intervals that each display variable ranges of GR values. Such irregular GR profiles are typically associated with rapidly fluctuating depositional environments of interbedded shales, siltstones and thin sandstones (downward GR spikes), occurring within an overall aggrading sequence. This facies typically represents neither a viable, conventional petroleum reservoir nor seal. There are several depositional environments that give rise to such GR profiles. For example, fluvial flood plains, tidal flats, onshore and offshore debris flows, shallow- and deep-water canyon fills. The thin sandstones tend not to be laterally extensive and lack substantial porosity and permeability. However, this class may be of interest in unconventional petroleum systems, as part of a viable petroleum-rich shale sequence. It is often important, when screening well logs, to distinguish such zones, particularly from class 2 shales.

2.2 GR feature augmentation

It is clearly apparent from the lithofacies features illustrated in Fig. 1 that a changing slope in the GR curve features in four of the characteristic features described (classes 1, 3, 5 and 6). On the other hand, two of the features (classes 2 and 4) are characterized by relatively flat and constant GR trends. Consequently, it is logical to consider GR derivatives as feature augments to use in the definition of the lithofacies classes described.

Eq. (1) defines the first derivative or rate of change in GR values ($d'GR$):

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

where GR_d is the GR value at depth d and GR_{d-1} is the GR value at depth $d-1$.

Eq. (2) defines the second derivative or rate of change in the rate of change (acceleration) in GR values ($d''GR_{dn}$):

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

where n represents a depth interval over which the second derivative $d''GR$ is calculated. The value of n can be varied to suit the lithological variations; applying higher n values for slow formation changes (i.e., thicker intervals for each facies), and applying lower n values for slow formation changes (i.e., thinner interbedded facies intervals). For this study an n value of 5 depth units is applied. The GR characteristics displayed by Classes 1, 3 and 5 (Fig. 1) show distinctive $d'GR$ and $d''GR$ values.

Another worthwhile GR derivative value to consider as a GR augmented characteristic is a simple moving average (SMA) of the first derivative, arithmetically averaged across an appropriate overlying depth interval. $SMA d'GR_{dn}$ is defined in Eq. (3):

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

For this study, the n value used for the overlying depth interval is 5 depth units, starting from the immediately preceding (shallower) depth interval to GR_d . This attribute can be used to overcome short-duration noise in GR depth curve and still respond to the underlying rate-of-GR-change trend. Most GR curves include frequent short-duration value blips, upwards and downwards responding to minor (thin) lithology anomalies; SMA values can help minimize their influence on the underlying GR trends. The greater the average depth interval range covered by such minor anomalies, the higher the n value applied in Eq. (3) should be.

Another set of GR attribute features worth considering are those used to define the volatility of the GR depth trend. Historical volatility is universally used in finance and economics when assessing price trend characteristics, specifically risk and the extent of historical fluctuations, of financial instruments, stock and options prices (Schwert, 1989; Glantz and Kissell, 2014; Danielsson et al., 2018). Used in its financial context, volatility (σ) represents the magnitude of variation of a publicly quoted (e.g., on a trading exchange) price series over a specified period of time, typically a number of trading days. It is commonly calculated as the standard deviation of natural logarithmic returns. This calculation can be usefully adapted to consider the historical volatility of a well log curve versus depth (instead of time). The first attribute to calculate is the natural logarithm (ln) of the instantaneous fluctuation in the GR value, as defined by Eq. (4):

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

where $i(d)$ represents the depth interval from $d-1$ to d .

The next step in the volatility (σ) calculation is to establish the standard deviation of the $\ln GR_{i(d)}$ values over a specified depth interval. For financial instruments the minimum time interval used in the σ calculation would typically be about 10 days but the number of trading days per month or quarter are also frequently used depending on the time horizon of interest. For well-log curve variations with depth, the depth interval selected should be meaningful in terms of the depth scale of the fluctuations observed in the log-curve trends. σ is calculated by applying Eq. (5).

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

where $\sigma_{i(dn)}$ is the volatility for the depth interval $d-1$ to $d-n$; L_{d-i} is the $\ln GR_{i(d)}$ for each overlying depth in the depth interval $d-1$ to $d-n$; and, $L_{id}mean$ is the arithmetic mean of all $\ln GR_{i(d)}$ values in the depth interval $d-1$ to $d-n$; and, n is the specified depth interval used. For this study an n value of 5 depth units is applied in the calculated volatility values. In practice, with different GR curves it is advisable to conduct sensitivity analysis to select the most appropriate n value to suit the dimensions of the lithofacies present. The GR characteristics displayed by Classes 2 and 4 versus Class 6 (Fig. 1) show distinctive $\sigma_{i(dn)}$ values.

As with the first derivative attribute (Eq. (3)) it is useful to also consider an SMA to calculate σ trends versus depth.

The SMA $\sigma_{i(dn)}$ is defined in Eq. (6).

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

For this study, the n value used for the overlying depth interval is 5 depth units, starting from the immediately preceding (shallower) depth interval to σ_d . Although, to demonstrate the technique with synthetic GR curves, the n values used in this study for Eqs. (2), (3), (5) and (6) are the same (i.e., 5 depth units) there is no reason why different n values could be used in each of these equations, as long as they are used consistently and are meaningful in terms of the scale of the facies variations encountered in a wellbore.

Hence, for the purpose of distinguishing lithofacies characteristics using only data derived from a GR curve, the following seven attributes for each depth point are evaluated as input variables (features):

- GR value (API units recorded),
- d' GR from Eq. (1),
- d'' GR $_{dn}$ from Eq. (2),
- SMA d' GR $_{dn}$ from Eq. (3),
- $\ln GR_{i(d)}$ from Eq. (4),
- $\sigma_{i(dn)}$ from Eq. (5),
- SMA $\sigma_{i(dn)}$ from Eq. (6).

2.3 Synthetic GR log curve versus depth to test extracted features

A synthetic GR curve has been created to display the six clastic lithofacies shapes described in Fig. 1. It has been artificially constructed to illustrate a range of GR responses that are characteristic of clastic sequences. It consists of 2250 GR data points and is displayed with depth units (meters) on the vertical scale in Fig. 2(a). The lithofacies classes assigned to specific depth intervals of the synthetic GR log are displayed in Fig. 2(b). Note that the GR curve shapes within each lithofacies are not smooth but include occasional anomalies (“blips”), and in some depth intervals gradient changes, to make them reflect more closely real GR curves recorded in wellbore logs. The upper 1450 depth points (0 to 1449, Fig. 2(a)) are used for training and validation of the ML models evaluated. The trained models are subsequently applied to predict the assigned facies classes for the lower 800 depth points (1450 to 2249, Fig. 2(a)) as an independent test. Both training/validation and testing subsets incorporate sections assigned facies classes 1 to 6.

The six GR feature augmentations (derivatives and volatility variables) described in section 2.4 are calculated for each of the 2250 GR data points for the synthetic GR curve (Fig. 2). The distributions of GR, its six calculated features and the facies classes are statistically summarized in Table 1.

Table 2 displays the Pearson correlation coefficients (R), in the form of a heat matrix, for the seven input variables and the facies classification scale (Class 1 to 6). The volatility related variables show moderate positive R values (~0.4) with the facies class number, whereas GR and GR derivative variables show poor negative R values (-0.14 to -0.8) with

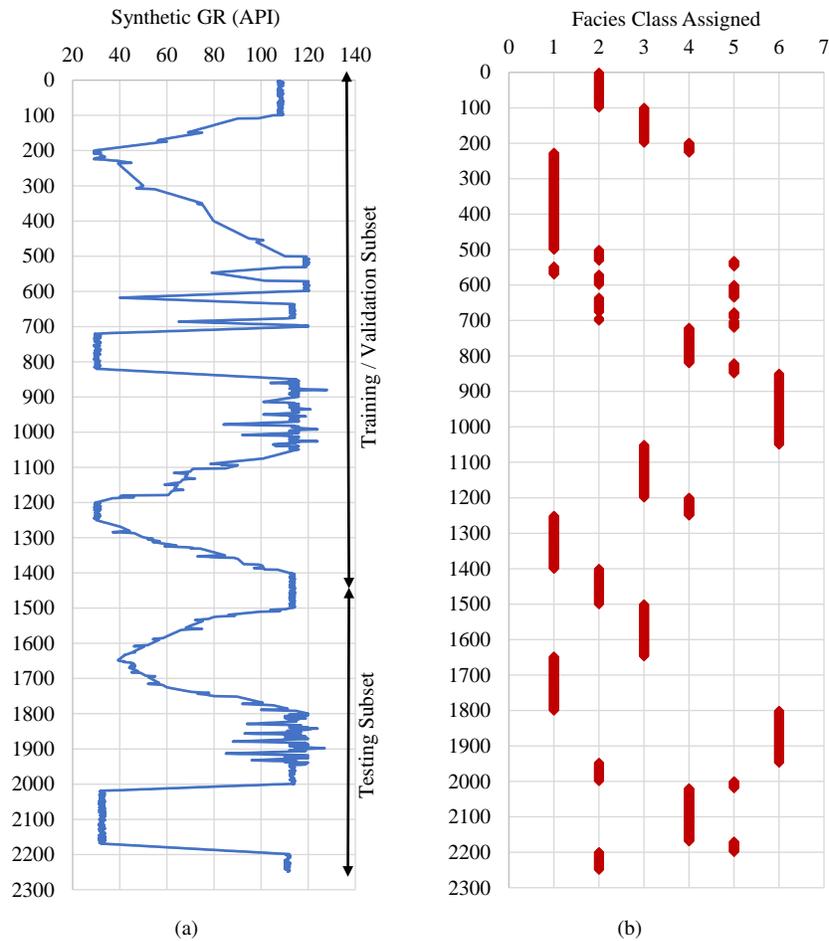


Fig. 2. (a) Synthetic GR curve versus depth; and (b) facies classes 1 to 6, as illustrated in Fig. 1 allocated to specific depth intervals in meters.

Table 1. Synthetic GR log variable distributions.

Statistical summary of seven GR attribute-variable and facies distributions for synthetic GR log								
2250 Data points	GR	$d'GR$	$SMA d'GR$	$d''GR_{dn}$	$\ln GR_{i(d)}$	$\sigma_{i(d)}$	$SMA \sigma_{i(dn)}$	Facies class
Minimum	29	-19.224	-5.000	-1.000	-0.382	0.00000	0.00000	1
5 th Percentile	31	-4.000	-2.400	-0.480	-0.050	0.00001	0.00002	1
10 th Percentile	32	-2.085	-1.200	-0.240	-0.035	0.00004	0.00005	1
25 th Percentile	48	-0.905	-0.358	-0.072	-0.012	0.00049	0.00327	1
50 th Percentile	84	0.100	0.052	0.010	0.001	0.01152	0.01265	3
75 th Percentile	111	0.825	0.395	0.079	0.010	0.02966	0.02960	4
90 th Percentile	114	2.342	1.200	0.240	0.035	0.04902	0.04559	6
95 th Percentile	117	4.000	2.600	0.520	0.054	0.05638	0.05271	6
Maximum	128	16.220	6.400	1.280	0.148	0.18112	0.17139	6
Average	79.3	0.002	0.002	0.000	0.000	0.01845	0.01847	3.0
Standard deviation	31.6	2.266	1.360	0.272	0.031	0.02051	0.01822	1.8
Standard error	0.666	0.048	0.029	0.006	0.001	0.000	0.000	0.037
Coefficient of variation	0.398	1227	785	781	1819	1.112	0.987	0.577

Table 2. Synthetic GR log variable distributions.

Correlation matrix (heat map) between seven gamma ray attributes and facies class for synthetic log								
R = Pearson correlation coefficient	GR	d' GR	SMA d' GR	d'' GR $_{dn}$	\ln GR $_{i(d)}$	$\sigma_{i(d)}$	SMA $\sigma_{i(dn)}$	Facies class
GR	1	0.0366	0.1268	0.1088	0.0328	-0.1996	-0.2194	0.1050
d' GR	0.0366	1	0.2520	0.5591	0.9020	0.0820	0.0172	-0.0824
SMA d' GR	0.1268	0.2520	1	0.8976	0.2668	0.0135	0.0784	-0.1413
d'' GR $_{dn}$	0.1088	0.5591	0.8976	1	0.5324	0.0591	0.0708	-0.1405
\ln GR $_{i(d)}$	0.0328	0.9020	0.2668	0.5324	1	0.1007	0.0426	-0.0797
$\sigma_{i(d)}$	-0.1996	0.0820	0.0135	0.0591	0.1007	1	0.6811	0.3883
SMA $\sigma_{i(dn)}$	-0.2194	0.0172	0.0784	0.0708	0.0426	0.6811	1	0.4231
Facies	0.1050	-0.0824	-0.1413	-0.1405	-0.0797	0.3883	0.4231	1

the facies class number. The contrasting relationships of the input variables with each other and with facies class suggest that the GR attributes should complement the GR values in distinguishing the facies characteristics defined in Fig. 1.

2.4 Machine learning models evaluated

Seven ML classification models are applied to evaluate the synthetic GR well log dataset (Table 1, Fig. 2). It is necessary to establish effective training and validation subset divisions of the data records. The training/validation split was established using the k -fold cross-validation technique (SciKit Learn, 2021a) and trial-and-error analysis. That analysis established that a random split of 80% of the data records to the training subset and the remaining 20% of data records to the validation subset lead to reliable and repeatable results with each ML model evaluated. The ML methods applied to the synthetic GR dataset belong to different types of algorithms. These models can all be configured to perform supervised learning based on training and validation subsets. Once successfully trained and validated, the models can subsequently be applied to independent testing data sets not previously seen by the models as part of their training and validation. The ML models belong to four distinct methodologies.

2.4.1 Tree-based ML models

The tree-based models employed are:

- Decision Tree (DT),
- Random Forest (RF, an ensemble decision tree),
- Adaboost or Adaptive Boosting (ADA; a boosted decision tree),
- Extreme Gradient Boosting (XGB; a boosted decision tree).

DT is a simple tree-based models in which data records are continuously split and assigned to different branches of the tree. Those branches are configured with nodes and leaves. The extent to which the tree is branched is controlled by the depth constraint.

ADA, RF, and XGB are more complex adaptations using ensembles, or architectures involving multiple trees. RF constructs multiple independent decisions trees with the train-

ing data, applying a bagging/out-of-bag technique based on variable distributions. Invariably the forecasts derived from the combined forest of DTs is of greater accuracy and more representative of the dataset as a whole than forecasts generated by individual DTs. ADA (Adaptive Boosting) re-assigns weights to each DT solution with each training iteration, focusing high weights on the data records responsible for the largest prediction errors in the previous iterations. XGB applies gradient boosting to an ensemble of DTs. The XGB algorithm is designed to deliver fast computational speeds and multiple tuning controls to make it as flexible as possible to suit a wide range of datasets.

2.4.2 Support vector classifier

The Support Vector Classifier (SVC) is a non-probabilistic binary linear classifier. The algorithm maps the training data records in multidimensional hyperspace space such that it maximises the distance between classes. It establishes the optimum position of a linear boundary (hyperplane) in multi-dimensional hyperspace, applying a kernel function to map lower dimensional data points into higher dimensions. The data records situated close to the optimum hyperplane are referred to as the “support vectors”, because they essentially define the position of the hyperplane.

2.4.3 Data-matching model

K-nearest Neighbor (KNN) is a non-parametric data-matching technique. Its training routine evaluates all the available training subset data records to establish and rank how closely they match each other, taking all input variables into account. KNN applies a similarity metric based on the magnitude of the collective errors of the input variables. The error measure is a distance function (typically Euclidian or Manhattan) with the closest matching records displaying the minimum error values. KNN predictions are based on the weighted values of a user-specified number of the best matching data records.

2.4.4 Multi-layer Perceptron

The Multi-layer Perceptron (MLP) is a widely used shallow neural network. MLPs involve a fully connected, feed-forward

Table 3. Hyperparameters values applied to seven ML algorithms configured to lithofacies classification based on a synthetic GR log.

Machine learning algorithms	Control parameter values applied
ADA	Number of estimators = 500; learning rate = 0.05; base estimator is DT with depth = 5; splitter = best
DT	Maximum depth = 1000; splitter = best
KNN	Neighbours (K) = 5; weighted by Manhattan distance
Multi-layer Perceptron (MLP)	3 hidden layers with 100, 50 and 25 neurons; activation fn. = relu; Solver = adam; alpha = 0.00005 maximum iterations 500; adaptive learning rate
RF	Number of estimators = 750; maximum depth = 150
SVC	Kernel = rbf; C = 10; gamma = 1.0
XGB	Number of estimators = 2000; maximum depth = 7; eta = 0.01; subsample = 0.7; columns sampled per tree = 0.5

SNN with an input layer, an output layer separated by one or more hidden layers. Each of the hidden layers is constructed with multiple nodes to which weights and biases are applied. Moreover, activation functions are applied when transferring information from one hidden layer to the next or to the output layer. MLP's are trained using back-propagation or other algorithms which adjust the weights and biases of the hidden layers. Through a series of iterations the back-propagation algorithm strives to minimize the errors in the collective predictions generated for the training subset data records. For the synthetic GR dataset, the MLP employs three fully-connected hidden layers.

2.4.5 ML model configurations

The mentioned ML models applied to classify lithofacies characteristics in the synthetic GR dataset are customized configurations in Python code of published packages (SciKit Learn, 2021b). As input variables to the ML models the seven GR and GR attribute variable values are all normalized. This avoids scale biases influencing the performance of the models. Normalization was configured to scale each variable in a range of -1 to $+1$ by applying Eq. (7).

$$Normx_i^m = \frac{2(x_i^m - x_{\min}^m)}{x_{\max}^m - x_{\min}^m} - 1 \quad (7)$$

where $Normx_i^m$ is the normalized variable value, x_i^m is the actual recorded value of the m^{th} variable of i^{th} data record, x_{\min}^m is the minimum value of the m^{th} variable considering all filtered data records and x_{\max}^m is the maximum value of the m^{th} variable considering all filtered data records.

It is unnecessary to describe here the underlying algorithms of these ML models as the calculation methodologies are well described in the literature. However, the ML model configurations, in terms of their controlling hyperparameter values and model architectures applied, need to be tailored to suit the synthetic GR dataset. That tailoring requires sufficient testing to justify the hyperparameter values applied. The optimized hyperparameters applied to each ML model are provided in Table 3. These values are in part determined by trial-and-error

analysis combined with Bayesian optimization (SciKit Learn, 2021c) and K-fold cross validation. The convergence trends of the algorithms were observed in multiple runs of the models to verify the effectiveness and repeatability of each model based on the optimal hyperparameter values (Table 3).

The ML models considered have been previously evaluated and mathematically described, in conjunction with well log/lithofacies prediction datasets with a range of dependent variables. Examples of the published studies relevant to lithofacies predictions from well-log-derived input variables for each model evaluated include: ADA (Farzi and Bolandi, 2016; Wrona et al., 2018); DT and KNN (Wrona et al., 2018; Merembayev et al., 2021); MLP (Mandal and Rezzaee, 2019; Hossein et al., 2020); RF (Hossein et al., 2020; Merembayev et al., 2021); SVR (Mandal and Rezzaee, 2019; Hossein et al., 2020); and, XGB (Bestagini et al., 2017; Shashank and Mahapatra, 2018). The mentioned studies apply and compare the performance of several ML models. Similarly, the evaluations presented here apply multiple ML models to evaluate the synthetic GR dataset involving GR rate of change and volatility variables.

2.5 Measures of prediction performance

Three statistical prediction performance measures are used to assess the results of the ML algorithms applied to the synthetic GR dataset. These metrics are:

Root Mean Squared Error (RMSE)

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2 \right]^{\frac{1}{2}} \quad (8)$$

where X_i is actual/measured facies class value for data record i ; Y_i is predicted facies class value for data record i in the subset being considered; n is total number of data records being predicted.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_i - Y_i| \quad (9)$$

Coefficient of Determination (R^2)

$$R^2 = \left[\frac{\sum_{i=1}^n (X_i - X_{mean})(Y_i - Y_{mean})}{\sqrt{\sum_{i=1}^n (X_i - X_{mean})^2 \sum_{i=1}^n (Y_i - Y_{mean})^2}} \right]^2 \quad (10)$$

The RMSE and MAE values, Eqs. (8) and (9) are expressed in terms of the facies class number scale of 1 to 6, not in normalized terms. Expressing the class prediction errors in this way makes it possible to readily assess their magnitude in terms of the range of facies class values encountered in the synthetic GR dataset.

Other important performance metrics for classification problems assess the absolute number and type of prediction errors, i.e., those data records for which an ML model predicts the facies class incorrectly. These measures are:

Absolute number of prediction errors recorded.

Accuracy (A) as determined by Eq. (11):

$$A = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (11)$$

where T_P = true number of positive predictions, T_N = true number of negative predictions, F_P = number of false positive predictions, and F_N = number of false negative predictions.

Precision (P) as determined by Eq. (12):

$$P = \frac{T_P}{T_P + F_P} \quad (12)$$

Recall (R) as determined by Eq. (13):

$$R = \frac{T_P}{T_P + F_N} \quad (13)$$

Balanced F-score (F_1), the harmonic mean of P and R , as determined by Eq. (14):

$$F_1 = \frac{2R}{R + P} \quad (14)$$

These classification accuracy measures are most usefully

displayed and interpreted in conjunction with an annotated confusion matrix.

3. Results

3.1 Formulaic distinction of lithofacies characteristics

Considering the GR attributes associated with the synthetic GR dataset, simple rules (formulas) can be empirically established by trial and error to distinguish, to a reasonable degree of accuracy, certain of the lithofacies classes (Fig. 2):

- **Class 1** (coarsening upwards) can be distinguished from other facies classes in cases where $0.01 \leq d'GR < 2.0$ or $0 < SMA d'GR < 2.0$ and $GR < 105$,
- **Class 2** (thick homogeneous shale) can be distinguished from other facies classes in cases where $GR \geq 105$ and $SMA \sigma_{i(dn)} < 0.02$,
- **Class 3** (fining upwards) can be distinguished from other facies classes in cases where $-2.0 < d'GR < -0.01$ or $-2 < SMA d'GR < 0$ and $GR < 105$,
- **Class 4** (massive homogeneous sandstone) can be distinguished from other facies classes in cases where $GR \leq 35$ and $-0.11 < d''GR_{dn} < 0.11$,
- **Class 5** (rapid transition zones) can be distinguished from other facies classes in cases where $d'GR \geq 2$ or $d'GR \leq -2$ and $35 < GR < 105$,
- **Class 6** (saw-tooth erratically interbedded shales and thin sandstones and silts) is more difficult to distinguish with a simple formula. An over oversimplistic approach can be adopted, i.e., to assign all data records not assigned to Classes 1 to 5 by the preceding formula to Class 6.

Applying these formulas results in 1813 correct classifications and 437 incorrect (19.4%) classifications for the 2250 data records of the synthetic dataset. The overall error statistics for this approach are RMSE = 1.323, MAE = 0.523, and $R^2 = 0.4898$. A more comprehensive analysis of the errors and

Formulaic Classification (2250 Data records)							
2250	Formulas	600	410	395	329	168	348
Total	Class	1	2	3	4	5	6
436	1	432	0	2	2	0	0
553	2	33	390	23	0	18	89
397	3	51	0	325	17	0	4
318	4	9	0	0	308	1	0
284	5	57	0	42	0	144	41
262	6	18	20	3	2	5	214
Precision:		72.00%	95.12%	82.28%	93.62%	85.71%	61.49%
437	Errors	168	20	70	21	24	134
80.58% Successful Predictions							
Class Prediction Performance:							
Accuracy		0.9134	0.9083	0.9274	0.9832	0.9170	0.9088
Precision		0.7200	0.9512	0.8228	0.9362	0.8571	0.6149
Recall		0.9908	0.7052	0.8186	0.9686	0.5070	0.8168
F1 Score		0.8340	0.8100	0.8207	0.9521	0.6372	0.7016

Fig. 3. Annotated confusion matrix providing error and accuracy analysis for each facies class applying the formulaic classification system defined in the text to all 2250 data records constituting the synthetic GR dataset displayed in Fig. 2.

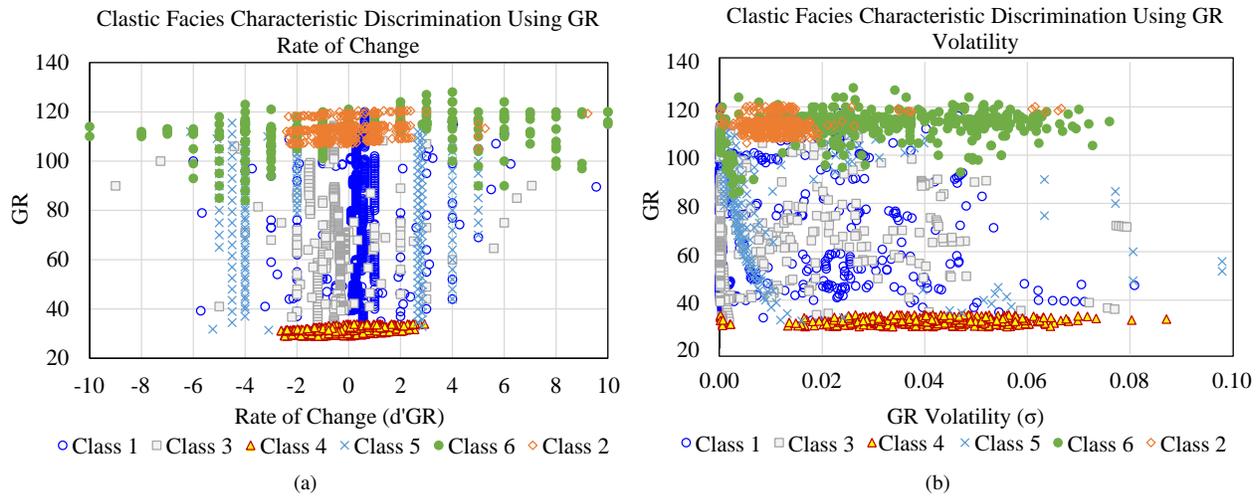


Fig. 4. Cross plots of GR versus: (a) GR rate of change, and (b) GR volatility for the 2250 data records constituting the synthetic GR dataset displayed in Fig. 2.

accuracy is provided a confusion matrix (Fig. 3).

The formulas perform a relatively good job at distinguishing records of Classes 2 and 4, as recorded by the precision scores in Fig. 3. However, Class 2 has a relatively low F_1 Score (0.81) because a substantial number of data records from other classes are wrongly assigned to Class 2. The simple formulas also correctly identify most of the data records in Classes 1, 3 and 5. However, a number of the data records in Classes 1 and 3, particularly those associated with the “blips” and near the boundaries of those classes, are confused mainly with Classes 5 and 2, but also between Classes 1 and 3 (Fig. 3). Class 5 achieves relatively high precision but has a low F_1 Score, because a substantial number of data records from other classes are wrongly assigned to Class 5. There are many erroneous predictions for the Class 6 data records; sections of Class 6 are confused primarily with Classes 2 and 5. Consequently, Class 6 achieves low precision and F_1 Score with the formulaic classification.

Notwithstanding the mentioned error count, the 80% success rate in predicting lithofacies using these simple formulas is an encouraging result. By writing more complex formulas, it is possible to bring the error count down below two hundred. However, for quick-look screening purposes these formula can be useful for identifying specific lithofacies from a gamma-ray log, e.g., during real time as an MWD log is recorded.

The simple formulas presented here to distinguish classes 1 to 6 are empirical and established by trial and error. The objective is to keep them as simple as possible so that they can be used as quick-look rules of thumb. Clearly, those formulas relate specifically to the GR dataset evaluated and would likely need modifications when applied to other datasets. It would be possible to generate a more complex set of formulas than those derived empirically using decision trees and other classification techniques. However, it is considered more beneficial to apply a range of machine learning classification methods that can be executed rapidly (i.e., in a few seconds) than developing more complex classification formulas that would take time and

need to be modified with each dataset evaluated.

3.2 Graphical distinction of facies characteristics based on GR attributes

Plotting GR versus GR rate of change ($d'GR$) (Fig. 4(a)) and GR volatility ($\sigma_{i(dn)}$) (Fig. 4(b)) highlights the value of these two GR attributes in distinguishing certain facies characteristics.

The GR values can themselves easily distinguish facies Class 2 from Class 4. The main benefit of the rate of change attribute is that for most data records distinguishes between Classes 1, 3 and 5. The bulk of Class 1 data records plot to the right of $d'GR = 0$ in Fig. 4(a), whereas the bulk of Class 3 data records plot to the right of $d'GR = 0$ in that figure, with both classes including some anomalous data records plotting in the region dominated by the other class. Moreover, the Class 5 data records are concentrated in regions further from $d'GR = 0$ in Fig. 4(a), making it possible to distinguish those rapid transition zones as well as their directions of slope. Classes 2 and 6 are not well separated in Fig. 4(a). However, they are slightly easier to distinguish in Fig. 4(b), albeit with substantial overlap in the low volatility region.

When the synthetic GR data records of each facies class are considered as groups they represent distributions with quite distinctive characteristics. Fig. 5 illustrates this by plotting mean GR for each class versus the coefficient of variation (CV) for two selected GR attributes: $1/SMAd'GR$ and $SMA\sigma_{i(dn)}$.

The CV is calculated by dividing the standard deviation of a distribution by its mean. A higher CV value typically reflects a more substantial degree of dispersion about the mean value of the class distribution. Each facies class plots in a distinctive area for the two GR attributes considered in Fig. 5, as they do for the other attributes not displayed. The distributions for the entire dataset (“All”) also plot in distinctive areas in Fig. 5. In order to plot the two attribute distributions on a similar horizontal scale the reciprocal of the CV of $SMAd'GR$ attribute is plotted in Fig. 5. For $SMA\sigma_{i(dn)}$

it is clear from their high CV values that the Class1 and Class 3 distribution show the greatest dispersion of values about their means, whereas Classes 2, 4 and 6 show the least dispersion about their means, and Class 5 and the entire dataset show intermediate dispersions about their means. For CV of $1/SMA d'GR$, as it is the reciprocal plotted, the classes plotting closest to zero on the horizontal scale are those displaying the greatest dispersion about their mean values. For this attribute it is the entire dataset and Class 6 that show the greatest dispersion about their means, Classes 1 and 3 show the least dispersion about their means, and Classes 2, 4, and 5 show intermediate dispersions about their means. The distinct dispersions of each of the class distributions is something that ML algorithms can usefully exploit.

Certain ratios are able to discriminate between the facies classes of the synthetic GR dataset to an extent. For instance, the ratio of the moving average rate of change to the moving average volatility ($SMA d'GR / SMA \sigma_{i(dn)} * 1000$), when plotted versus the GR value (Fig. 6(a)), is able to separate several

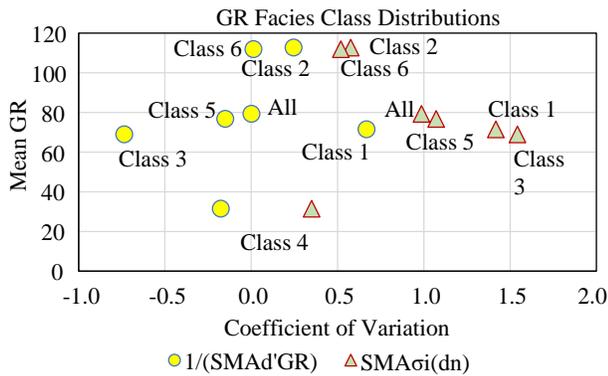


Fig. 5. Cross plot of mean GR versus the reciprocal of GR moving average rate of change, and the GR moving average volatility for the six GR facies classes and entire dataset (All) distinguished in the synthetic GR dataset displayed in Fig. 2.

of the facies classes. In fact, the ratio of volatility attributes ($\ln GR_{i(d)} / \sigma_{i(dn)}$) does a slightly better job at discriminating between Classes 1, 3 and 5 (Fig. 6(b)), particularly the different steep sloping directions of the Class 5. On the other hand, a ratio of the first to second GR derivatives ($d'GR / d''GR$) plotted versus GR (Fig. 7) values does not provide useful segregation of Classes 1, 3, and 5.

Considering collectively the GR attribute characteristics and relationships displayed in Figs. 4 to 7, it is apparent that the GR derivative and volatility variables used together provide more useful discrimination between the clastic facies distinguished in the synthetic GR dataset. This justifies their inclusions as input feature for ML analysis.

3.3 ML analysis of clastic facies using GR derivative and volatility variables

The strategy adopted for applying the seven ML models to the synthetic GR dataset was to initially justify using the 80%:20% split of data records in the training/validation dataset (data record numbers 0 to 1449, Fig. 2). This is followed by comparing of the facies class prediction accuracy achieved by the training and validation subsets and applying the trained models to all 1450 data records. The trained models are then applied to the independent testing subset (data record numbers 1450 to 2249, Fig. 2). The model performances are ranked and the results of the best performing model are displayed and evaluated using enhanced confusion matrices.

3.3.1 K-fold cross-validation results

Analysis of each ML model with the data records 0 to 1449 was performed with several different k-folds and different combinations of the GR attributes as a feature selection comparison to predict the facies classes for each data record. 5-fold cross validation produced the most consistent results (Table 4) and its results are shown for the following five feature combinations:

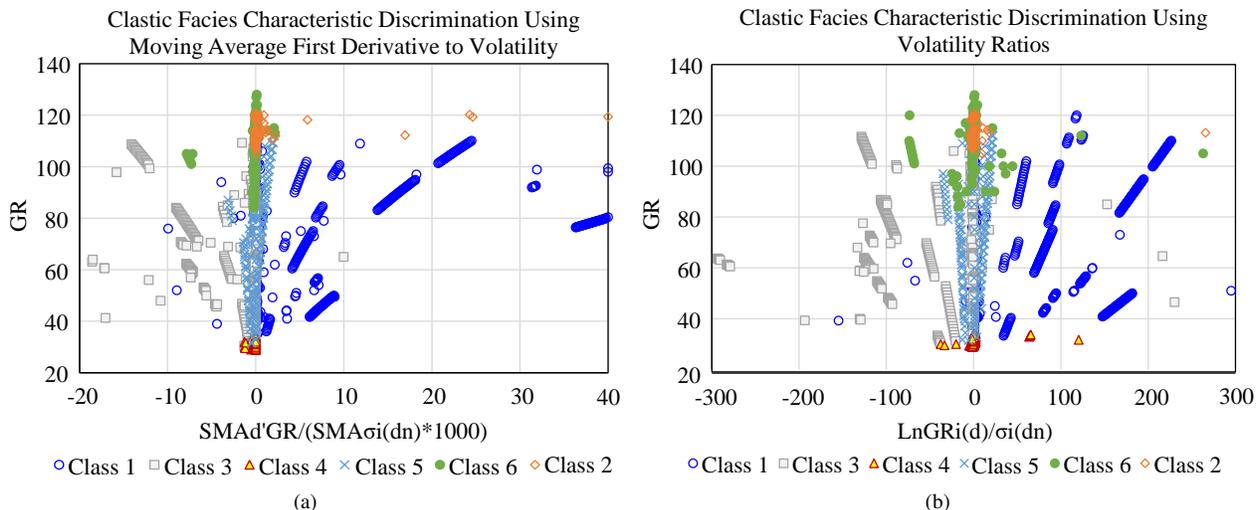
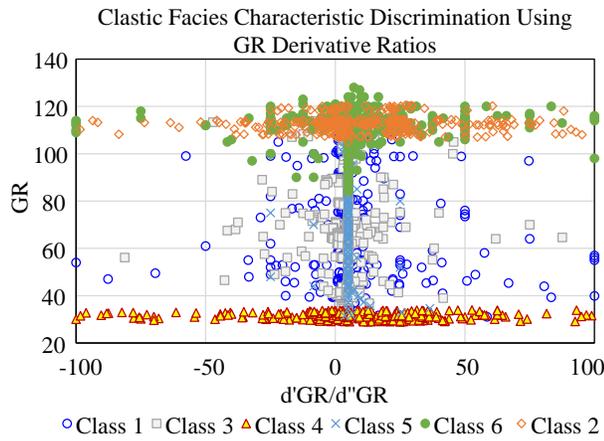


Fig. 6. Cross plots with facies classes distinguished of: (a) GR versus the ratio of the moving average rate of change to the moving average volatility; and, (b) GR versus the ratio of volatility attributes.

Table 4. Results of 5-fold cross-validation based on 15 cases. σ^2 MAE refers to standard deviation of the MAE values considering all 15 cases run for each model.

5-Fold cross-validation results for ML models applied to data records 0 to 1449 (15 cases run for each model)										
Model	7 Input variables		6 Input variables		5 Input variables		4 Input variables		3 Input variables	
	MAE	σ^2 MAE								
ADA	0.1864	0.0467	0.1759	0.0449	0.1903	0.0491	0.2120	0.0499	0.2894	0.0581
DT	0.2607	0.0376	0.2589	0.0441	0.2722	0.0646	0.2754	0.0357	0.3572	0.0430
KNN	0.3113	0.0362	0.3175	0.0322	0.3297	0.0497	0.3322	0.0496	0.4011	0.0654
MLP	0.2506	0.0479	0.2802	0.0678	0.2274	0.0516	0.3133	0.0633	0.3949	0.0574
RF	0.1637	0.0429	0.1775	0.0461	0.1722	0.0513	0.1945	0.0577	0.2972	0.0504
SVC	0.3011	0.0484	0.2982	0.0536	0.3324	0.0586	0.4025	0.0548	0.4671	0.0625
XGB	0.1048	0.0237	0.1159	0.0316	0.1039	0.0296	0.1156	0.0378	0.2545	0.0640

**Fig. 7.** Cross plot of GR versus the ratio of first to second GR derivatives ($d'GR/d''GR$) with facies classes distinguished. Classes 1, 3, and 5 are not as well separated as for the ratios displayed in Fig. 6.

- 7-variable data: GR, $d'GR$, $SMA d'GR$, $d''GR$, $\ln GR_{i(d)}$, $\sigma_{i(dn)}$, $SMA \sigma_{i(dn)}$,
- 6-variable data: GR, $d'GR$, $SMA d'GR$, $d''GR$, $\sigma_{i(dn)}$, $SMA \sigma_{i(dn)}$,
- 5-variable data: GR, $d'GR$, $SMA d'GR$, $\sigma_{i(dn)}$, $SMA \sigma_{i(dn)}$,
- 5-variable data: GR, $d'GR$, $SMA d'GR$, $SMA \sigma_{i(dn)}$,
- 4-variable data: GR, $SMA d'GR$, $SMA \sigma_{i(dn)}$.

Although these variable combinations may not be optimal, particularly for those combinations with six or less variables, they identify that even using very few of the available variables can result in relatively high prediction performance for the facies classes.

Each 5-fold cross-validation execution was executed three times, resulting in 15 cases, for which the mean MAE and the standard deviation (σ^2 MAE) were calculated. The σ^2 MAE values range from 0.0237 (XGB 7-variable model) to 0.678 (MLP 6-variable model) which are low (with reference to the facies class scale of 1 to 6) indicating high reproducibility of all the models and feature selections shown in Table 4. The lowest mean MAE achieved by the 5-fold analysis is 0.1039

for XGB (5-variable model) with models XGB, RF and ADA, in that order generating the best prediction accuracy in terms of MAE for each feature selection. On the other hand, the DT, MLP, KNN and SVC models generated higher mean MAE values, with the performance of KNN and SVC deteriorating more than the other models for the 3- and 4-variable feature selections. Overall, the Table 4 results suggest that the 7-variable feature selection models outperform the models with fewer features, confirming that all seven variables are being used effectively by the models to improve their facies class prediction performance.

A more rigorous feature selection sensitivity analysis was conducted applying each of the trained and validated ML models to the unseen testing subset (data records 1451 to 2250) using different combinations of variables. This analysis revealed that removing any one of the variables from the 7-variable model resulted in poorer facies class prediction performance for the testing subset than that achieved by the 7-variable model for five of the seven ML models. In particular, for the three best performing models (RF, XGB and SVR), the 7-variable model outperformed any other combination of variables. The feature selection analysis therefore indicates that the 7-variable model offers the best combination of features.

3.3.2 Training and validation of the ML models

Encouraged by the results of Table 4, the ML models were trained and validated for each feature selection configuration evaluated by k -fold cross-validation. The results for the 80:20 data record splits (1160 records for training and 290 for validation) are displayed for selected high-performing ML models in Table 5, together with the results for the trained model applied to all 1450 data records (the "All" case).

Fig. 8 compares the ML model results for the validation data set in terms of MAE versus number of erroneous data record classifications.

The trained ML models display a wide range of classification accuracies with the validation subset. Overall, the 5-, 6- and 7-variable models tend to outperform 3- and 4-variable

Table 5. Selected ML model training and validation classification results applied to the first 1450 data records of the synthetic GR dataset using 80% training: 20% validation splits. The training and validation results for all ML models evaluated are shown in a supplementary file.

ML training and validation results for clastic facies prediction					
RF	7-Var	6-Var	5-Var	4-Var	3-Var
Train R ²	1.0000	1.0000	1.0000	1.0000	1.0000
Train RMSE	0.0000	0.0000	0.0000	0.0000	0.0000
Train MAE	0.0000	0.0000	0.0000	0.0000	0.0000
Validation R ²	0.7790	0.8070	0.7472	0.7472	0.6605
Validation RMSE	0.7900	0.7381	0.8449	0.8449	0.9791
Validation MAE	0.2172	0.2000	0.2448	0.2586	0.3103
Total error number	28	27	30	34	35
All R ²	0.9597	0.9648	0.9539	0.9539	0.9381
All RMSE	0.3533	0.3301	0.3778	0.3778	0.4379
All MAE	0.0434	0.0400	0.0490	0.0517	0.0621
SVC					
Train R ²	0.7680	0.7631	0.7054	0.6476	0.5651
Train RMSE	0.8570	0.8660	0.9658	1.0562	1.1733
Train MAE	0.2259	0.2345	0.2793	0.3483	0.4284
Validation R ²	0.6544	0.6153	0.5982	0.4773	0.4724
Validation RMSE	0.9879	1.0422	1.0651	1.2149	1.2205
Validation MAE	0.3414	0.3759	0.3897	0.4828	0.4759
Total error number	43	46	49	56	55
All R ²	0.7474	0.7362	0.6859	0.6167	0.5484
All RMSE	0.8847	0.9040	0.9865	1.0898	1.1829
All MAE	0.2490	0.2628	0.3014	0.3752	0.4379

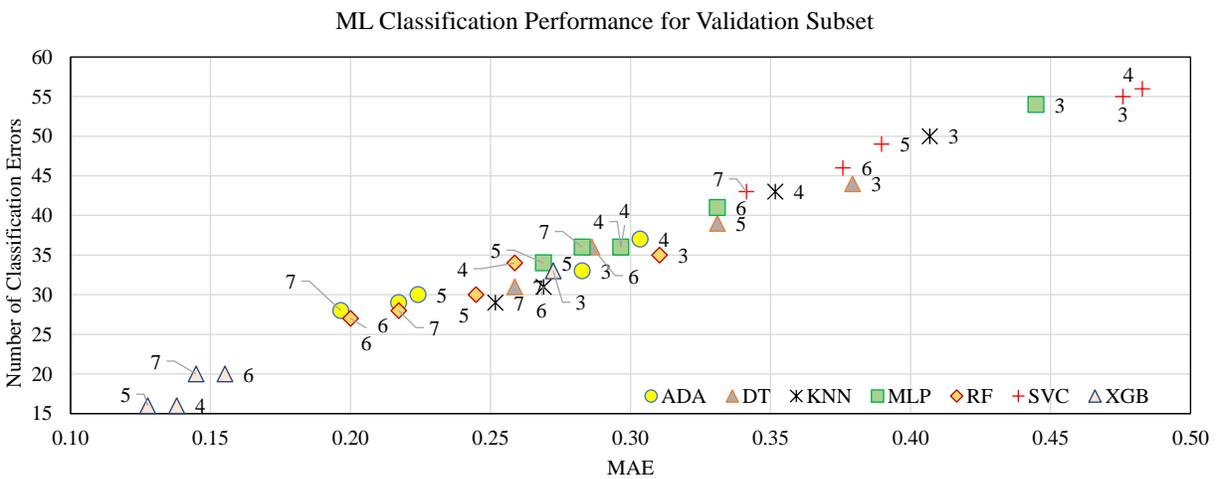


Fig. 8. Cross plot of classification errors versus MAE for seven ML models applied to the validation subset of the synthetic GR dataset. The number label refer to the number of input variables in the models (see text), for example, 7 means seven input variables are involved.

Table 6. Selected ML trained model classification results for the unseen testing subset results applied to the last 800 data records of the synthetic GR dataset. The testing subset results for all ML models evaluated are shown in a supplementary file.

ML unseen testing subset results for clastic facies prediction					
RF	7-Var	6-Var	5-Var	4-Var	3-Var
Test R ²	0.7355	0.7030	0.6566	0.6347	0.4449
Test RMSE	0.8853	0.9381	1.0087	1.0404	1.2826
Test MAE	0.2513	0.2700	0.3075	0.3425	0.5300
Total error number	82	81	91	109	165
Execution time (seconds)	23.2	25.2	23.5	21.9	21.4
SVC					
Test R ²	0.7305	0.7237	0.7073	0.6398	0.6140
Test RMSE	0.8937	0.9048	0.9314	1.0332	1.0695
Test MAE	0.2713	0.2763	0.2925	0.3500	0.3938
Total error number	95	96	102	119	138
Execution time (seconds)	5.0	5.7	5.0	5.1	4.9

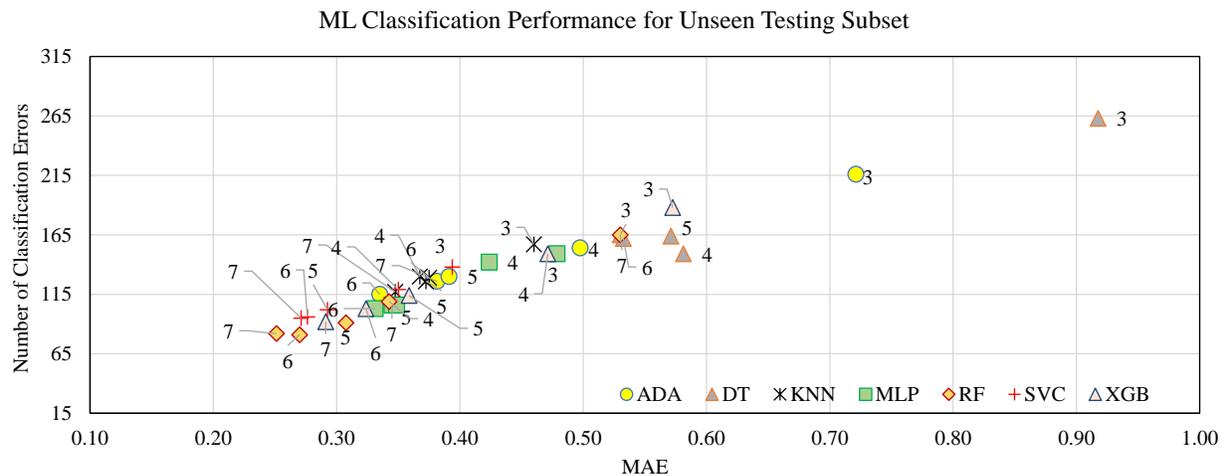


Fig. 9. Cross plot of classification errors versus MAE for seven ML models applied to the unseen testing subset of the synthetic GR dataset. The number label refer to the number of input variables in the models (see text).

models for each ML. However, the 7-variable model is not always the highest performing (e.g., XGB and RF). The XGB models achieve the lowest number of errors and lowest MAE values, some distance ahead of the RF and ADA models in Fig. 8. On the other hand, the SVC model generates the highest number of errors and highest MAE values for all of its models applied to the validation subset. However, the near perfect fits generated for the training subset and the much lower accuracies for the validation subset (Table 5) indicates that several of the models have, to an extent, overfitted the training dataset. Nevertheless the results of Table 5 are consistent with those of the 5-fold cross validation analysis (Table 4).

3.3.3 Applying the trained ML models to an unseen testing sequence

A key test for the reliability of the ML models is applying the trained models to unseen independent data. The trained

models have therefore been applied to the independent testing subset (data records 1451 to 2250) with the classification results displayed in Table 6 and Fig. 9.

It is notable that the RF, SVC and XGB trained models in that order, generate the most accurate results with the unseen testing subset. A comparison of Figs. 8 and 9 suggests that the XGB models have overfitted the training subset to a greater extent than the other models, whereas the SVC model has done so to a lesser extent than the other models. Again, the 5-, 6- and 7-variable models tend to outperform 3- and 4-variable models for each ML algorithm applied to the testing subset. Indeed, the 3- and 4-variable models tend to generate substantially less accuracy than the other models. SVC generates the best results compared to the other 3-variable models, and the spread of results between the 7-variable and 3-variable models is least for SVC (Fig. 9). The DT model generates the poorest classification results for the

Expanded Confusion Matrices for Independent Unseen Testing Subset (Data records Numbers 1450 to 2249)															
RF 7-Variable Model								SVC 3-Variable Model							
800	RF	153	149	135	149	53	161	800	RF	174	174	105	150	62	135
Total	Class	1	2	3	4	5	6	Total	Class	1	2	3	4	5	6
153	1	128	1	15	0	1	8	153	1	120	9	14	0	1	9
153	2	0	146	0	0	0	7	153	2	1	147	0	0	0	5
147	3	21	1	119	0	0	6	147	3	48	6	88	0	3	2
150	4	1	0	1	148	0	0	150	4	0	0	0	149	1	0
50	5	2	0	0	1	42	5	50	5	5	2	1	1	40	1
147	6	1	1	0	0	10	135	147	6	0	10	2	0	17	118
Precision:		83.66%	97.99%	88.15%	99.33%	79.25%	83.85%	Precision:		68.97%	84.48%	83.81%	99.33%	64.52%	87.41%
Errors		25	3	16	1	11	26	Errors		54	27	17	1	22	17
89.75% Successful Predictions								82.75% Successful Predictions							
Class Prediction Performance:															
Accuracy		0.9349	0.9863	0.9423	0.9958	0.9742	0.9497	Accuracy		0.8838	0.9525	0.8970	0.9970	0.9539	0.9350
Precision		0.8366	0.9799	0.8815	0.9933	0.7925	0.8385	Precision		0.6897	0.8448	0.8381	0.9933	0.6452	0.8741
Recall		0.8366	0.9542	0.8095	0.9867	0.8400	0.9184	Recall		0.7843	0.9608	0.5986	0.9933	0.8000	0.8027
F1 Score		0.8366	0.9669	0.8440	0.9900	0.8155	0.8766	F1 Score		0.7339	0.8991	0.6984	0.9933	0.7143	0.8369

Fig. 10. Annotated confusion matrices providing error and accuracy analysis for each facies class applying the trained RF 7-variable and SVC 3-variable models to the last 800 data records of the synthetic GR dataset displayed in Fig. 2.

testing subset, by some distance.

The execution times for the models are listed in Table 6. All models were run on a laptop computer with a dual core processor of 1.8 GHz with a maximum memory speed of 2.2 GHz. The execution times represent the inclusive time taken to perform 5-*k* cross validation, test and validate the models and evaluate the unseen testing subset. The DT, KNN and SVC models operate the fastest, taking about 5 seconds to complete. The ADA and RF models take between 20 and 30 seconds to complete, whereas the XGB (~80 seconds) and MLP (~100 seconds) models are slowest. The fast operating speeds mean that it is feasible to run an ensemble of models and compare their performances. Based on a comparison of the results for the testing and validation subsets running an ensemble of models seems to be an advisable strategy.

Overall, the RF 7-variable model achieves the best results with the unseen testing subset. Fig. 10 displays an annotated confusion matrix for the classification results of that model together with the high performing SVC 3-variable model.

Both models displayed in the Fig. 10 confusion matrices are highly successful at classifying facies classes 2 and 4, as shown by their high accuracy and precision values. However, the RF 7-variable model performs much better than the SVC 3-variable model in correctly classifying facies classes 1, 3 and 5. F1 Scores for those classes are > 0.8 for the RF model but around 0.7 for the SVC model. Both models perform quite well at classifying Class 6 (F₁ Scores > 0.8), substantially better than the formulaic approach (Fig. 3). For the RF model its main problem is confusing 21 Class 1 data records as Class 3 and 15 Class 3 data records as Class 1.

Fig. 11 provides a point by point comparison of actual facies class with predicted facies class for the best-performing RF 7-variable model applied to the unseen testing subset. The cluster of erroneous predictions around the 1650 data record level, confusing Class 1 with Class 3 and Class 3 with Class 1, make a substantial contribution to the overall errors associated with this model. A close inspection of Fig. 2 reveals that this zone of errors is located around the low-GR apex of the symmetrical feature comprising of a coarsening upwards fea-

ture turning abruptly into a fining-upwards feature. The errors occur at low GR values where GR rates of change are low but oscillating from positive to negative. This abrupt change from Class 1 to Class 3 is not present in the training/validation sequence (Fig. 2).

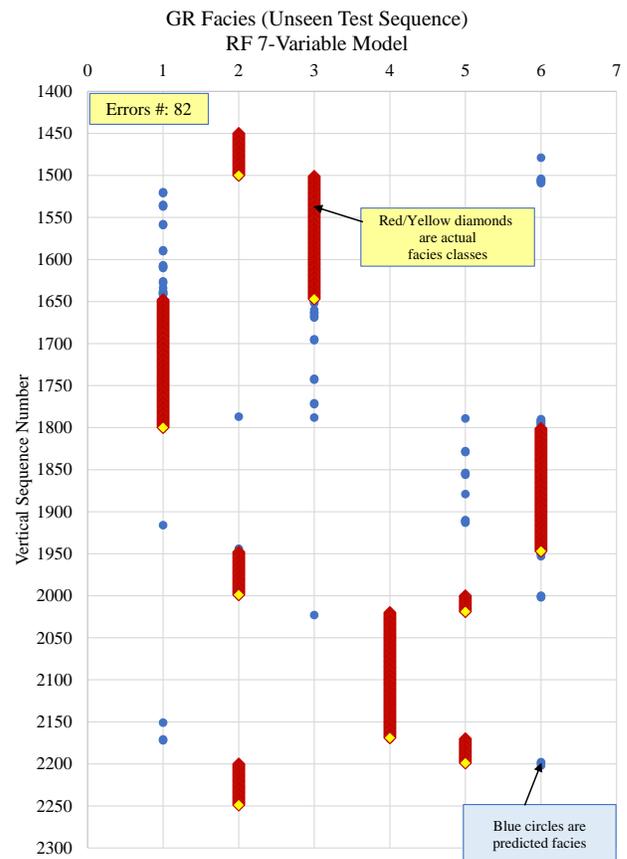


Fig. 11. actual versus predicted facies classes for the unseen testing subset of the synthetic GR dataset achieved by the RF 7-variable model. Where predictions are correct they are overlain by the actual values (red/yellow diamonds), so only the erroneous predictions are revealed by the blue dots.

The results for the best performing model highlight that, in order to work effectively, the ML models need to be trained using sequences that include representative examples of facies features that occur in the unseen sections to be classified. The fact that the 5-, 6- and 7-variable ML models are the most accurate clearly indicates that the ML models are effectively using most of the derivative and volatility attribute variables with which they are provided as inputs to improve their classification accuracy and precision.

4. Discussion

The results of the study demonstrate the value of extracting GR derivative and volatility attributes to characterize certain clastic sequence facies characteristics using only GR data as the input. It is not intended to suggest that GR be used in isolation to provide comprehensive definitive facies analysis, which clearly requires additional mineralogical, lithological and bio-stratigraphy information. However, there are certain circumstances, where provisional facies analysis derived from the GR log can be useful and effective. These circumstances are:

- Quick-look, real-time analysis as well logs are recorded. By applying the formulaic and/or ML methodologies proposed to measurement-while-drilling (MWD) and wire-line logs as they are being recorded can quickly and reliably reveal the presence of certain facies characteristics,
- In exploration, screening thousands of meters of GR data from a wellbore to identify certain features known to be associated with high quality oil/gas reservoirs in a basins (e.g., coarsening upward sections) is an important task. Such analysis can be achieved rapidly with the aid of the GR derivative and volatility attributes defined in this study,
- For many wells drilled, cost and time constraints mean that only very basic logging suites are recorded (e.g., GR/sonic/resistivity). In some wells only the GR log is recorded over some of the shallower sections above target zones. Moreover, when more comprehensive well logging suites are run, or those including spectral gamma ray, shear velocity, etc., they are often only recorded over limited depth intervals. This leaves large sections of wellbores poorly logged, from which useful facies information could be extracted using the GR attribute analysis proposed,
- For many older wellbores, only limited and poor resolution logging suites are available making it impossible to apply more sophisticated log-based facies analysis to those sections,
- Many well-log based facies classification techniques relying on multiple suites of well log curves tend to overlook the valuable shape information contained within the GR curve. Combining GR derivatives and volatility attributes together with other well logs (e.g., sonic, density, neutron etc.) has the potential to improve the prediction accuracy of such models.

This study is designed to introduce and demonstrate the value of using GR derivative and volatility attributes for clastic

sequence facies analysis. The results presented justify further work to evaluate the full potential of the method. Such future research planned includes:

- Applying the proposed techniques to real GR logs recorded in multiple wellbore sections to test the generalization of the method across a developed oil/gas field,
- Evaluating the potential to extract derivative and volatility attributes from other well logs, such as compressional sonic logs, which are also widely available and recorded in most wellbores, for facies feature analysis. Such attributes could potentially be used in conjunction with the GR attributes to improve clastic facies predictions (e.g., to reduce some of the confusion between Class 2 and Class 6 experienced in the dataset evaluated in this study),
- The ML models evaluated are fast to execute but are not particularly transparent in the details they provide regarding how each of their specific data record predictions are made. For detailed data mining purposes, and to better explore the subtleties of complex facies more transparent algorithms are required. Optimized data matching algorithms (e.g., transparent open box, Wood, 2018), specifically configured for data mining facies predictions, offer potential to improve the facies insight provided by the proposed GR-attribute technique,
- Evaluating the potential to exploit derivative and volatility attributes of GR (and other logs) to predict certain facies characteristics in carbonates and other non-clastic sequences.

5. Conclusions

Extracting detailed shape attributes from GR log curves provides effective variables with which to reliably identifying certain clastic sequence facies characteristics. This study demonstrates that rate of change ($d'GR$), second derivative ($d''GR$) and volatility (σGR), when used together and in conjunction with GR values can distinguish and classify six commonly occurring facies characteristics in clastic sequences (i.e., coarsening upward, transgression, fining upward, aggradation, rapid transition and sawtooth features). Moving averages over specified depth intervals, tailored to suit the sections, help to further refine the usefulness of the GR derivative and volatility attributes for facies analysis. These attributes can be used formulaically or as inputs to ML models to effectively classify the clastic facies in a synthetic GR sequence divided into training, validation and unseen testing subsets. 5-fold cross validation models demonstrate the repeatability of the ML models using an 80%:20% split between training and validation. ML models configured with between three and seven input variables (GR values plus three derivative attributes and three volatility components) demonstrate the models with 5-, 6- and 7-variables generate the best facies classifications with the validation and testing sequences. A random forest model with seven input variables outperforms the other models and formulaic classification. That model generates 82 erroneous classifications from 800 data records in the unseen testing sequence. It also achieves a mean absolute error of 0.2513 and a root mean squared error of 0.8853 (in relation to a clastic

facies class scale of 1 to 6) for that testing sequence. The fact that the RF model uses all seven input variables to generate the best classification highlights the value of combining both derivative and volatility variables, including moving averages, as input features. Annotated confusion matrices, configured specifically for this study, offer a useful visual technique for concisely assessing and comparing the classification accuracy and precision of the multiple ML and formula models evaluated for facies classification purposes.

The proposed method offers the potential to conduct rapid screening of measurement while drilling and wireline logs for clastic facies features as those logs are recorded. Moreover, the results of this study suggest that derivative and volatility GR attributes should be further evaluated for use in conjunction with other well-log and/or core data to assist in more complex facies characteristics and analysis.

Acknowledgement

The author thanks the reviewers for their constructive comments.

Supplementary file

<https://doi.org/10.46690/ager.2022.01.06>

Conflict of interest

The author declares no competing interest.

Open Access This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- Bestagini, P., Lipari, V., Tubaro, S. A machine learning approach to facies classification using well logs. Paper SEG-2017-17729805 Presented at the 2017 SEG International Exposition and Annual Meeting, Houston, Texas, 24-29 September, 2017.
- Busch, J. M., Fortney, W. G., Berry, L. N. Determination of lithology from well logs by statistical analysis. *SPE Formation Evaluation*, 1987, 2(4): 412-418.
- Danielsson, J., Valenzuela, M., Zer, I. Learning from history: Volatility and financial crises. *The Review of Financial Studies*, 2018, 31(7): 2774-2805.
- Dubois, M. K., Bohling, G. C., Chakrabarti, S. Comparison of four approaches to a rock facies classification problem. *Computers & Geosciences*, 2007, 33(5): 599-617.
- Dypvik, H., Eriksen, D. Natural radioactivity of clastic sediments and the contributions of U, Th and K. *Journal of Petroleum Geology*, 1983, 5(4): 409-416.
- Emery, D., Myers, K. *Sequence Stratigraphy*. Blackwell Science, Oxford, UK, 1996.
- Fadokun, D. O., Oshilike, I. B., Onyekonwu, M. O. Supervised and unsupervised machine learning approach in facies prediction. Paper SPE-203726-MS Presented at the SPE Nigeria Annual International Conference and Exhibition, Virtual, 11-13 August, 2020.
- Faga, A. T., Oyenyin, B. M. Effects of diagenesis on neural-network grain-size prediction. Paper SPE-60305-MS Presented at the SPE Rocky Mountain Regional/Low-Permeability Reservoirs Symposium and Exhibition, Denver, Colorado, 12-15 March, 2000.
- Farzi, R., Bolandi, V. 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. *Modeling Earth Systems and Environment*, 2016, 2: 105.
- Glantz, M., Kissel, R. *Multi-Asset Risk Modeling: Techniques for a Global Economy in an Electronic and Algorithmic Trading Era*. Academic Press, New York, USA, 2014.
- Goncalves, C. A., Harvey, P. K., Lovell, M. A. Application of a multilayer neural network and statistical techniques in formation characterization. Paper SPWLA-1995-FF Presented at the SPWLA 36th Annual Logging Symposium, Paris, France, 26-29 June, 1995.
- Hall, B. Facies classification using machine learning. *The Lead Edge*, 2016, 35(10): 906-909.
- Halotel, J., Demyanov, V., Gardiner, A. Value of geologically derived features in machine learning facies classification. *Mathematical Geosciences*, 2020, 52: 5-29.
- Hossain, T. M., Watada, J., Aziz, I. A., et al. Machine learning in electrofacies classification and subsurface lithology interpretation: A rough set theory approach. *Applied Sciences*, 2020, 10(17): 5940.
- Hurst, A. Natural gamma-ray spectroscopy in hydrocarbon bearing sandstones from the Norwegian continental shelf. *Geological Society*, 1990, 48(1): 211-222.
- Ippolito, M., Ferguson, J., Jenson, F. Improving facies prediction by combining supervised and unsupervised learning methods. *Journal of Petroleum Science and Engineering*, 2021, 200: 108300.
- Krumbein, W. C., Sloss, L. L. *Stratigraphy and Sedimentation*. W H Freeman & Co, London, UK, 1951.
- Mandal, P. P., Rezaee, R. Facies classification with different machine learning algorithm—An efficient artificial intelligence technique for improved classification. *ASEG Extended Abstracts*, 2019, 1: 1-6.
- Merembayev, T., Kurmangaliyev, D., Bekbauov, B., et al. A Comparison of machine learning algorithms in predicting lithofacies: Case studies from Norway and Kazakhstan. *Energies*, 2021, 14(7): 1896.
- Reverdy, X., Argaud, M., Walgenwitz, F. Minerological analysis required for log interpretation in complex lithologies. *Transactions of the SPWLA 8th European Symposium*, 1983.
- Rider, M. H. *Geological Interpretation of Well Logs*. Blackie, New York, USA, 1986.
- Rider, M. H. Gamma-ray log shape used as a facies indicator: Critical analysis of an oversimplified methodology. *Geological Society*, 1990, 48(1): 27-37.
- Rogers, S. J., Fang, J., Karr, C., et al. Determination of lithology from well logs using a neural network. *AAPG Bulletin*, 1992, 76(5): 731-739.
- Russell, W. L. The total gamma ray activity of sedimentary rocks as indicated by Geiger counter determinations. *Geophysics*, 1944, 9(2): 180-216.

- Scholle, P. A., Spearing, D. Sandstone Depositional Environments. American Association of Petroleum Geologists, tulsa, UAS, 1982.
- Schwert, G. W. Why does stock market price volatility change over time. *The Journal of Finance*, 1989, 154 (5): 1115-1153.
- SciKit Learn. Cross-validation: evaluating estimator performance 2021a. <https://scikit-learn.org/stable/>
- SciKit Learn. Bayesian optimization of hyperparameters in Python 2021c.
- Selley, R. C. Concepts and Methods of Subsurface Facies Analysis. American Association of Petroleum Geologists Education Course Notes, Tulsa, USA, 1978.
- Shashank, S., Mahapatra, M. P. Boosting rock facies prediction: weighted ensemble of machine learning classifiers. Paper SPE-192930-MS Presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 12-15 November, 2018.
- Tran, T. V., Ngo, H. H., Hoang, S. K., et al. Depositional facies prediction using artificial intelligence to improve reservoir characterization in a mature field of Nam Con Son Basin, Offshore Vietnam. Paper OTC-30086-MS Presented at the Offshore Technology Conference Asia, Kuala Lumpur, Malaysia, 17-19 August, 2020.
- Wood, D. A. A transparent Open-Box learning network provides insight to complex systems and a performance benchmark for more-opaque machine learning algorithms. *Advances in Geo-Energy Research*, 2018, 2(2): 148-162.
- Wood, D. A. Lithofacies and stratigraphy prediction methodology exploiting an optimized nearest-neighbour algorithm to mine well-log data. *Marine and Petroleum Geology*, 2019, 110: 347-367.
- Wood, D. A. Bakken stratigraphic and type well log learning network for transparent prediction and rigorous data mining. *Natural Resources Research*, 2020, 29(2): 1329-1349.
- Wrona, T., Pan, I., Gawthorpe, R. L., et al. Seismic facies analysis using machine learning. *Geophysics*, 2018, 83(5): O83-O95.