[Arti](https://doi.org/10.1016/j.aiig.2022.02.007)fi[cial Intelligence in Geosciences 2 (2021) 148](https://doi.org/10.1016/j.aiig.2022.02.007)–[164](https://doi.org/10.1016/j.aiig.2022.02.007)





Contents lists available at [ScienceDirect](http://www.sciencedirect.com/science/journal/26665441)

Artificial Intelligence in Geosciences

journal homepage: [www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences](http://www.keaipublishing.com/en/journals/artificial-intelligence-in-geosciences)

[](http://crossmark.crossref.org/dialog/?doi=10.1016/j.aiig.2022.02.007&domain=pdf)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*

A R T I C L E I N F O

*Keywords:*

*Rolling average derivatives Log-curve volatility Lithofacies log characteristics Confusion analysis*

*Gamma-ray attributes*

*Well-log feature augmentation.*

A B S T R A C T

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 pre- dictions. 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 at- tributes 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](#_bookmark79)). The possibility of exploiting recorded well-log data to assist in lithofacies classification dates back to the mid-twentieth century ([Russell, 1944](#_bookmark72); [Krumbein and Sloss, 1951](#_bookmark59)). At that time, the well logs available that were able to provide reasonable reso- lution 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](#_bookmark74) [Spearing, 1982](#_bookmark74); [Rogers et al., 1992](#_bookmark71); [Emery and Myers, 1996](#_bookmark35)).

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

([Rider, 1986](#_bookmark69), [1990](#_bookmark70); [Cant, 1992](#_bookmark31)). Additionally, the GR responses to grain size of clastic formations ([Hurst, 1990](#_bookmark53)) make it possible to distinguish important lithofacies features such as fining- or coarsening upwards ([Kesslar and Sachs, 1995](#_bookmark57); [Faga and Oyeneyin, 2000](#_bookmark37)). In carbonates, and other more varied and heterogeneous lithological sequences additional geological and mineralogical information is typically required to com- plement well-log data sufficiently to provide reliable lithofacies classifi- cation ([Reverdy et al., 1983](#_bookmark66); [Halotel et al., 2020](#_bookmark49)). 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 compre- hensive well-log suites over the entire wellbore sections drilled. Conse- quently, 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](mailto:dw@dwasolutions.com). <https://doi.org/10.1016/j.aiig.2022.02.007>

Received 1 February 2022; Received in revised form 27 February 2022; Accepted 27 February 2022

Available online 7 March 2022

2666-5441/© 2022 The Authors. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

methods ([Goncalves et al., 1995](#_bookmark46)), have been applied to well-log data to assist in lithofacies prediction since the 1980s ([Busch et al., 1987](#_bookmark30)). However, in multi-lithology sections well-log data typically displays non-linear relationships with lithofacies, making such methods unreli- able and relatively inaccurate. Various artificial-neural-network models, particularly multi-layer perceptrons (MLP), are extensively used to pre- dict facies classes from well-log data ([Rogers et al., 1992](#_bookmark71); [Puskarczyk,](#_bookmark64) [2019](#_bookmark64); [Hossein et al., 2020](#_bookmark51)). One attraction of MLP is that for complex

* + 1. *GR derivative attributes*

The first derivative (*d*’*GR*) is calculated with Eq. [(1)](#_bookmark0).

*d*' *GR* = (*GR* — *GR* ) / *Abs*(*d* — (*d* — 1)) (1)

*d*

*d*—1

where: *GRd* = recorded GR value at GR curve location *d* (depth in feet or meters); *GRd*—1 = recorded GR value at GR curve location *d-*1.

The second derivative (*d*’’*GR**dn*) is calculated with Eq. [(2)](#_bookmark1).

carbonate sequences they can generally outperform regression, proba- ' '

bilistic (e.g., Bayesian classification) and fuzzy classifier methods in lithofacies classification tasks using multiple well-log curves ([Dubois](#_bookmark34) [et al., 2007](#_bookmark34); [Agrawal et al., 2022](#_bookmark26)). Neural networks have also been usefully hybridized with a Markov transition matrix ([Feng et al., 2018](#_bookmark39); [Feng, 2020](#_bookmark40)) to form Bayesian neural networks ([Feng, 2021](#_bookmark41)) for lith- ofacies classification using well-log data.

*d*''*GRdn* = (*d GRd* — *d GRd*—*n*) / *Abs*(*d* — (*d* — *n*)) (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)](#_bookmark2).

In recent years, the high lithofacies prediction performance of ensemble and boosted-tree-based algorithms has become apparent ([Xie](#_bookmark89) [et al., 2019](#_bookmark89)). This is particularly the case for carbonate and complex

*SMAd*’*GRdn* =

*i*=*n*

*i*=1

X

'

*d*—*i*

*d GR*

!,*n* (3)

lithological sequences by several studies ([Hall, 2016](#_bookmark47); [Hall and Hall, 2017](#_bookmark48); [Al-Mudhafar, 2017](#_bookmark27); [Bestagini et al., 2017](#_bookmark28), [Shashank and Mahapatra,](#_bookmark80) [2018](#_bookmark80)). 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 classifica- tion ([Sun et al., 2020](#_bookmark81)). Additionally, regression-free K-nearest neighbor (KNN) ([Merembayev et al., 2021](#_bookmark61)) and the transparent open box algo- rithm [Wood (2019](#_bookmark85), [2020)](#_bookmark86) have been successfully applied to challenging lithofacies and stratigraphic classification and data mining using well-log data. Combinations of data-matching algorithms with clustering tech- niques have also been usefully applied ([Potratz et al., 2021](#_bookmark63)). Integrating ML models, clustering techniques ([Fadokun et al., 2020](#_bookmark36)), geological data ([Ippolito et al., 2021](#_bookmark55)) and bio-stratigraphic information ([Tran et al.,](#_bookmark82) [2020](#_bookmark82)) with well log data can also be effective for facies classification purposes, in cases where sufficient data is available.

Recently ([Wood, 2022](#_bookmark87)), 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

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 depending on the nature of the lithological section drilled. Finely interbedded strata with high volatility are likely to be better character- ized 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.

* + 1. *GR volatility attributes*

[Wood (2022)](#_bookmark87) 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 (*LnGRi*) repre- sents an intermediate step in establishing volatility. It is calculated with Eq. [(4)](#_bookmark3)

*LnGRi*(*d*) = *Ln*(*GRd* / *GRd*—1) (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 (σi) of *LnGRi*(*d*) for a specific sequence of GR data points. It is calculated with Eq. [(5)](#_bookmark4).

sﬃPﬃﬃﬃﬃﬃ*n*ﬃﬃﬃﬃﬃ(ﬃ*L*ﬃﬃﬃdﬃﬃﬃﬃiﬃﬃ—ﬃﬃﬃﬃﬃ*L*ﬃﬃﬃ*i*ﬃ*d*ﬃﬃ*m*ﬃﬃﬃ*e*ﬃﬃ*a*ﬃﬃﬃ*n*ﬃﬃ)ﬃﬃ2**ﬃ**

classification in multiple lithology sections with limited well-log coverage. A logged section though varied lithologies (clastic, carbon-

*σi*(*dn*) =

*i*=1 —

n — 1

(5)

ate, 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.

1. Method and materials

where: *σi*(*dn*) = interval volatility between depths *d-1* and *d-n*; *L*d—i = *LnGRi*(*d*) for each GR data point in the interval between depths *d-1* and *d- n*; *Lidmean* = arithmetic mean of *LnGRi*(*d*) between depths *d-1* and *d-n;n* = interval over which σi 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 (SMAσi) of the GR volatility attribute. It is calculated with Eq. [(6)](#_bookmark5).

X

* 1. *Well-log feature augmentation*

It has been recently demonstrated using a synthetic GR log that de-

*SMAσi*(*dn*) =

*i*=*n i*=1

*σd*—*i*

!,*n* (6)

rivative 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](#_bookmark87)). That method- ology 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.

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)](#_bookmark2)), 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)](#_bookmark4), if the trial and error analysis suggests it.

The three derivative GR attributes and the three volatility GR attri- butes defined by Eqs. [(1)](#_bookmark0)–[(6)](#_bookmark0) 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.

* 1. *Borehole logged section evaluated*

Triassic section penetrated in the Winterborne Kingston —1 (WK-1) well The logged wellbore section evaluated in this study is from the bore drilled in 1981 in the western part of the petroliferous Wessex Basin

in the Southern United Kingdom ([Rhys et al., 1982](#_bookmark68)). The section of in- terest 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 li- thologies 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](#_bookmark62) [et al., 2021](#_bookmark62)).

These feature of the WK-1 Triassic section pose a meaningful chal- lenge 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 success- fully developed and applied ([Wood, 2022](#_bookmark87)).

The Triassic stratigraphy of the Western Wessex basin is summarized in [Fig. 1](#_bookmark6).

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 Bran- scombe Mudstone Formation, a thin but distinctive (from its well-log characteristics) anhydrite and limestone zone exists at about 1865m depth. This is a representation of the Red Rock Gypsum Layer distributed sporadically in the Wessex Basin ([Hounslow and Ruffell, 2006](#_bookmark52)). The Blue Anchor formation is the uppermost formation of the Mercia Mudstone Group. Its grey/green colored dolomitic siltstones with sporadic anhy- drite/gypsum layers display well log characteristics ([Howard et al.,](#_bookmark54) [2008](#_bookmark54)), in terms of GR, DT and PB values, which are difficult to distin- guish 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](#_bookmark7) 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](#_bookmark7)), which are more gradational between some intervals (e.g., the top of the Sherwood Sandstone and base of the Mercia Mudstone) but quite

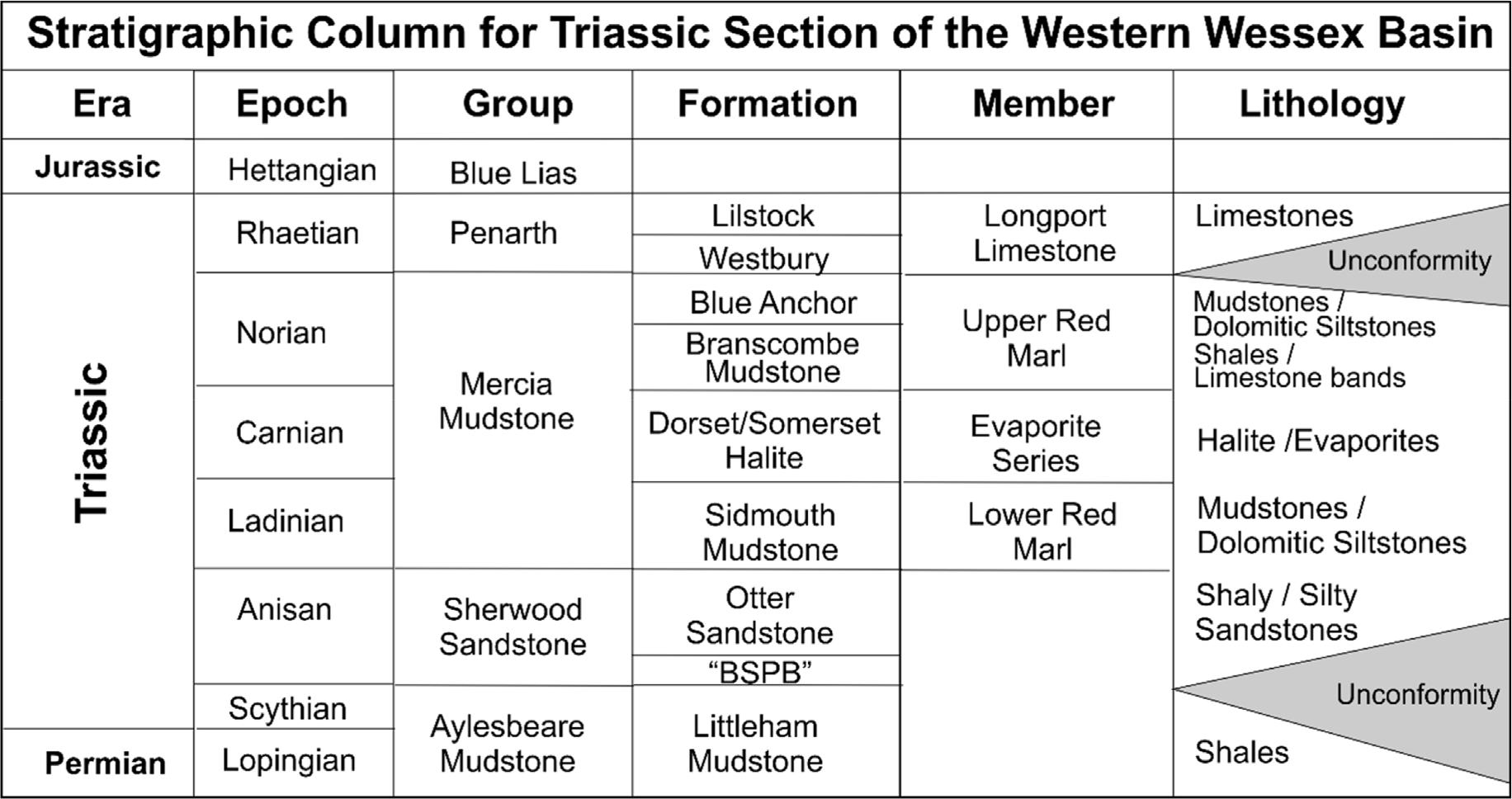


Fig. 1. Summary stratigraphy and lithology of the Wessex Basin Triassic section relevant to the Winterborne Kingston WK-1 wellbore location. Incorporating in- formation from [Underhill and Stoneley (1998)](#_bookmark83), [Hounslow and Ruffell (2006)](#_bookmark52); [Howard et al. (2008)](#_bookmark54); [Kaya (2015)](#_bookmark56). BSPB refers to the Budleigh Salterton Pebble Beds.

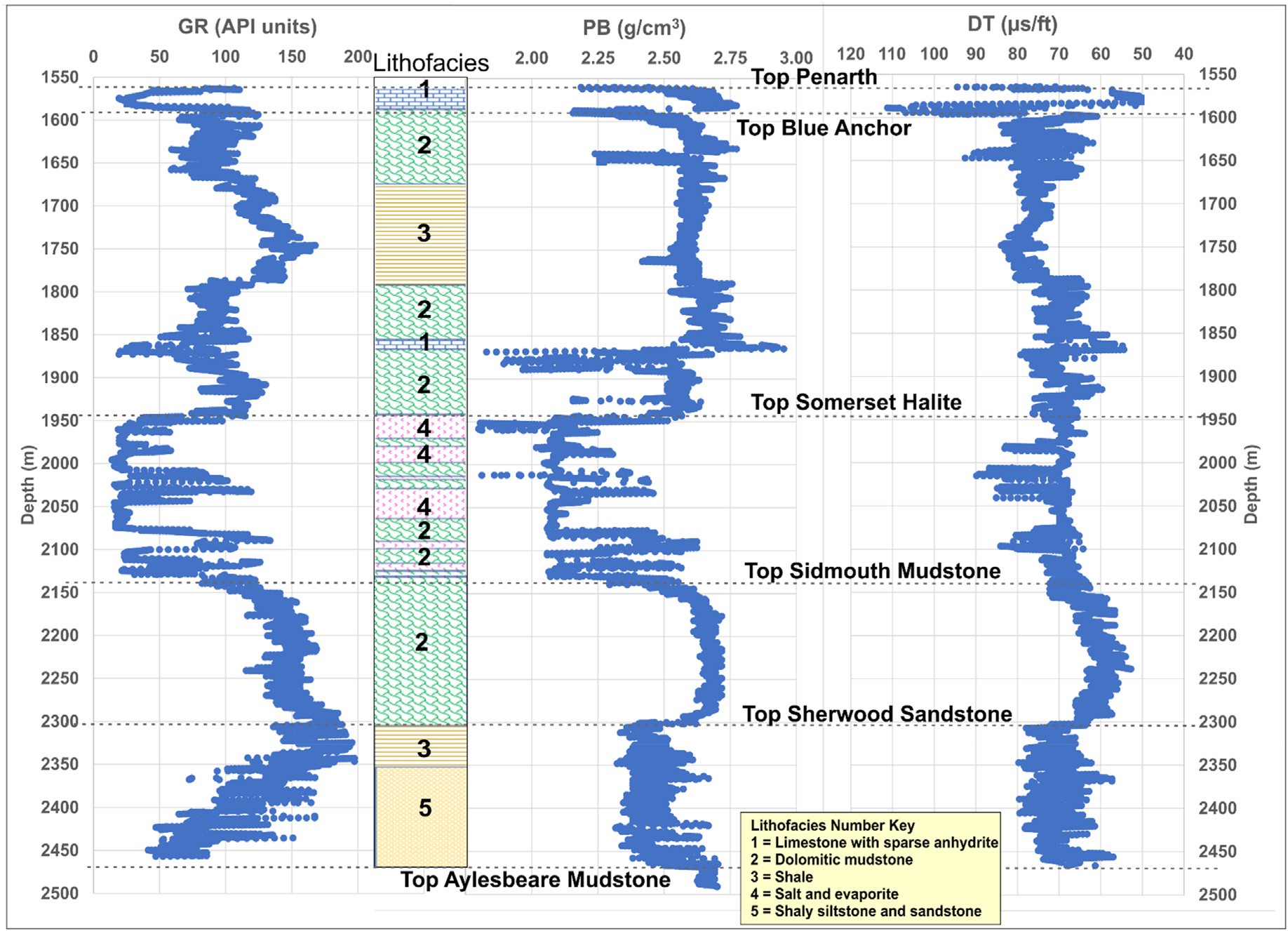


Fig. 2. GR, PB and DT well logs recorded for the Triassic section of the WK-1 wellbore with 5 distinct lithofacies sequences distinguished as classes 1 to 5.

abrupt between other intervals (e.g., the contact at the top of the Som- erset Halite with the overlying dolomitic mudstone). One of the advan- tages 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](#_bookmark8) 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](#_bookmark10) 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](#_bookmark11). 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](#_bookmark10)) 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](#_bookmark12) displays a heat map of the calculated correlation coefficient values between the log variables and facies class. The values for all col- umns in [Fig. 6](#_bookmark12) 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](#_bookmark12)) 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](#_bookmark12)). The R values between the GR attributes and facies class are substantially lower than

highest negative R value with facies class (—0.49), compared to —0.21 for GR and +0.21 for DT. those displayed by the GR, DT, PB recorded log data. PB displays the

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 repre- sentative 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 recor-

ded well log variables and/or, in particular the calculated GR-attribute variables.

* 1. *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](#_bookmark75) [Learn, 2022a](#_bookmark75)).

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

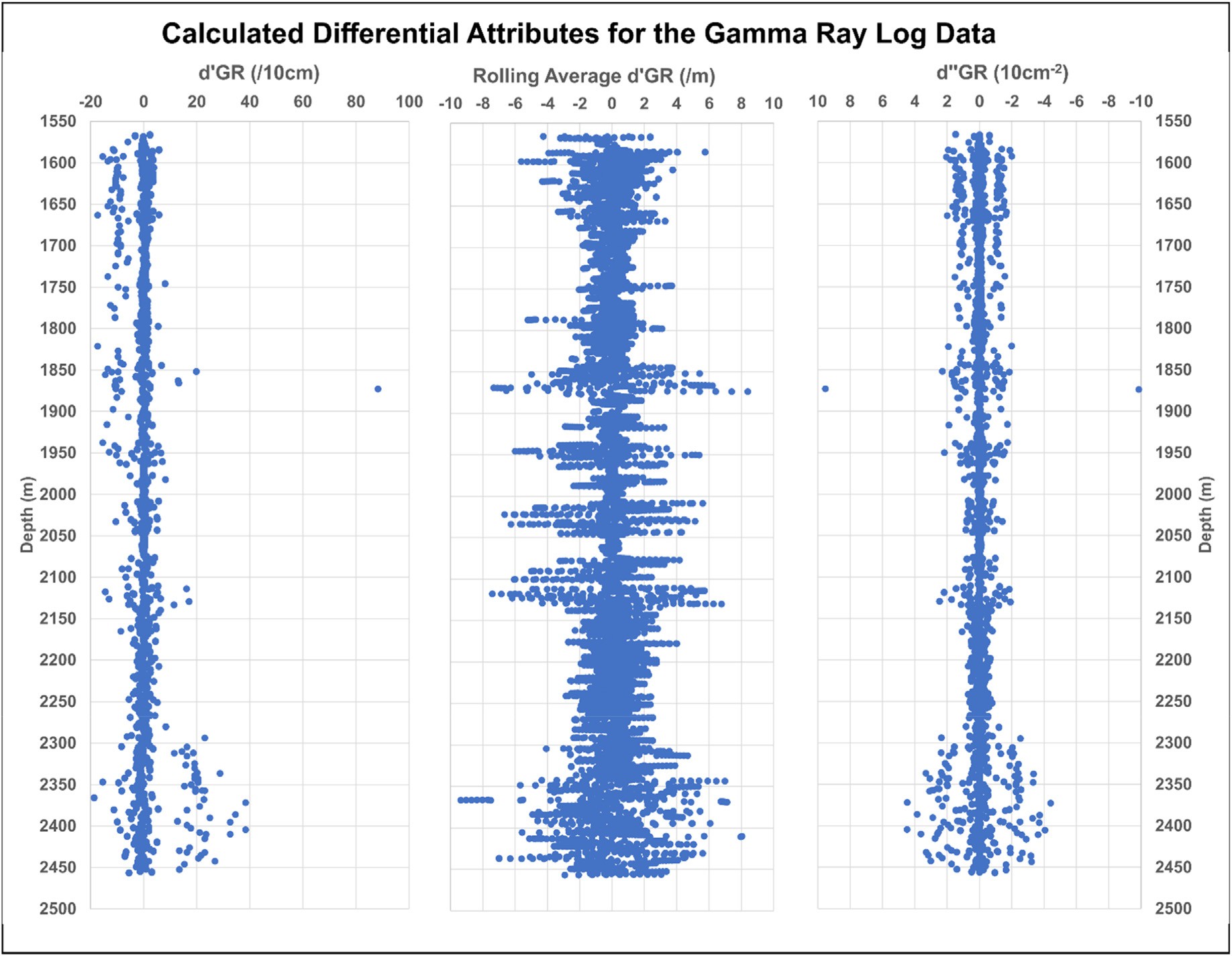


Fig. 3. Three GR derivative attributes calculated for each GR data point of the Triassic section of the WK-1 wellbore. These derivative attributes are the first derivative (d’GR), the moving average per meter of the first derivative (SMAd’GR) and the second derivative (d’‘GR).

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

ADA: Adaboost ([Freund and Schapire, 1997](#_bookmark43); [Wrona et al., 2018](#_bookmark88)). DT: Decision tree ([Quinlan, 1986](#_bookmark65); [Sarkar and Majundar, 2020](#_bookmark73)). KNN: K-nearest neighbor ([Fix and Hodges, 1951](#_bookmark42); [Merembayev et al.,](#_bookmark61)

[2021](#_bookmark61)).

MLP: Multi-layer perceptron ([Rosenblatt, 1958](#_bookmark67); [Hossein et al., 2020](#_bookmark51)). OLR: Ordinary least-squares regression ([Goldberger, 1964](#_bookmark45); [Gao,](#_bookmark44)

[2011](#_bookmark44)).

RF: Random forest ([Ho, 1998](#_bookmark50); [Kim et al., 2018](#_bookmark58); [Farzi and Bolandi,](#_bookmark38) [2016](#_bookmark38)).

SGD: Stochastic gradient descent multi-linear regression ([Bottou,](#_bookmark29) [1998](#_bookmark29); [Agrawal et al., 2022](#_bookmark26)).

SVC: Support vector classifier ([Cortes and Vapnik, 1995](#_bookmark33); [Mandal and](#_bookmark60) [Rezaee, 2019](#_bookmark60)).

XGB: Extreme gradient boosting ([Chen and Guestrin, 2016](#_bookmark32); [Bestagini](#_bookmark28)

and Bayesian optimization ([SciKit Learn, 2021b](#_bookmark77)) 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](#_bookmark13). Applying the optimum configurations and control parameters, the models can then be evaluated with K-fold cross valida- tion ([SciKit Learn, 2022b](#_bookmark78)) to establish their repeatability with various random splits of the dataset into training and testing subsets.

* 1. *Data preprocessing*

models are each normalized to scale range of —1 to +1. Such normali- The well log data and calculated GR attributes fed as input to the zation 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)](#_bookmark9).

[et al., 2017](#_bookmark28)).

*Normxm*

2\* *xm* — *xminm*

1 (7)

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

*i*

*i xmaxm* — *xminm*

=

—

where: *Normxm* = variable value adjusted to a —1 to +1 scale; xm = actual

*i* i

the list above), therefore their detailed methodologies are not described here. However, the models do require configuration and tuning adjust- ments 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](#_bookmark76))

bution; *xminm* = minimum value of *m*th variable distribution; *xmaxm* = recorded/calculated value for *i*th data point in the *m*th variables distri- maximum value of *m*th variable distribution.

* 1. *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

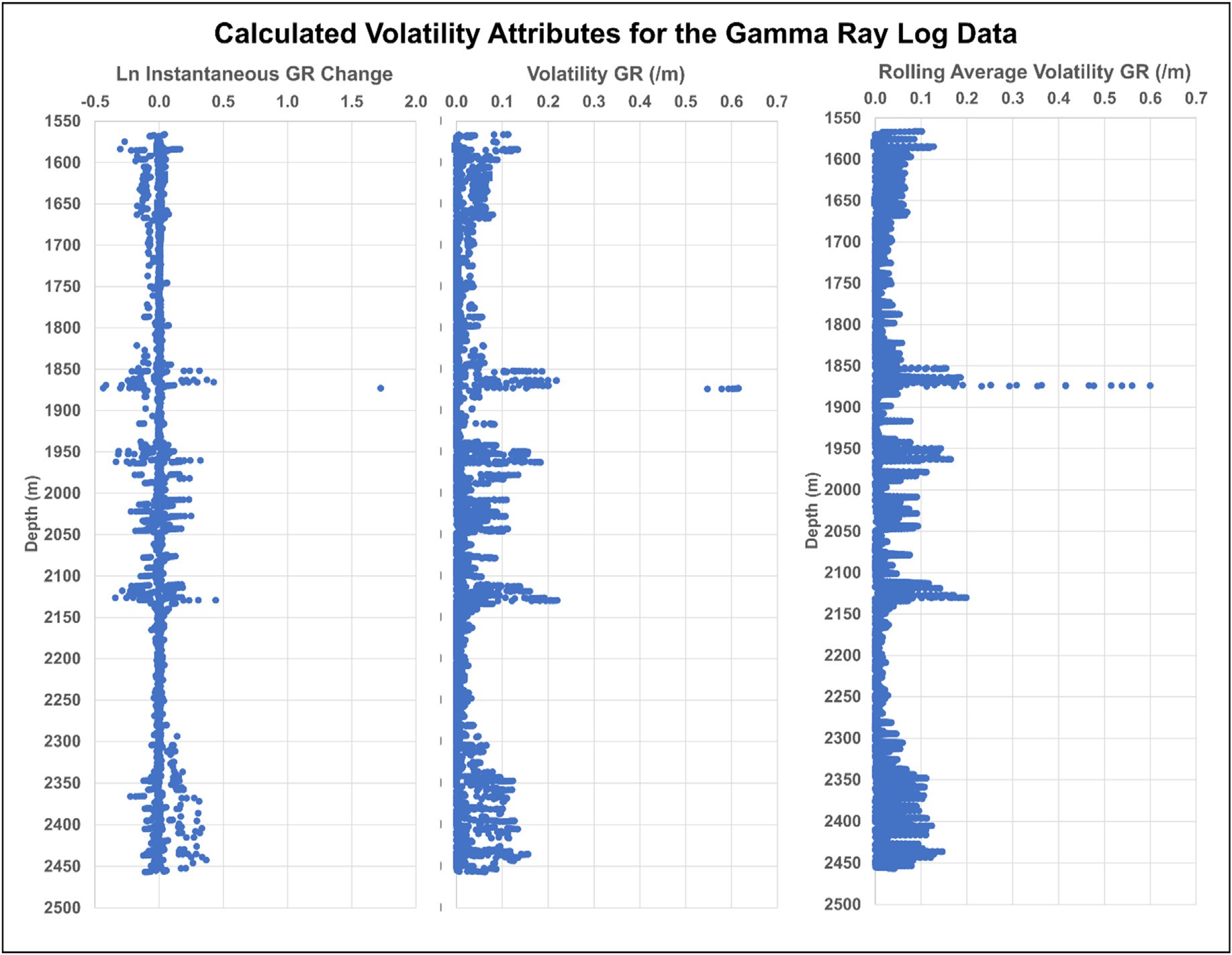


Fig. 4. Three GR volatility attributes calculated for each GR data point of the Triassic section of the WK-1 wellbore. These volatility attributes are the instantaneous change (LnGRi), volatility (σGR) and the moving average per meter of the volatility (SMAσGR).

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 | SMAd'GR | d''GRdn | LnGRi(d) | σi(dn) | SMAσi(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](#_bookmark22). These metrics are:

Root Mean Squared Error (RMSE) Mean Absolute Error (MAE) Coefficient of Determination (R2)

P

Total number of prediction errors ( Error) Accuracy (A)

Precision (P) Recall (R)

Balanced F-score (F1)

It is essential to consider the statistical measures of prediction per- formance ( Error, RMSE, MAE and R2) in conjunction with measures relating specifically to misclassifications (A, P, R, F1).

1. Results

P

* 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

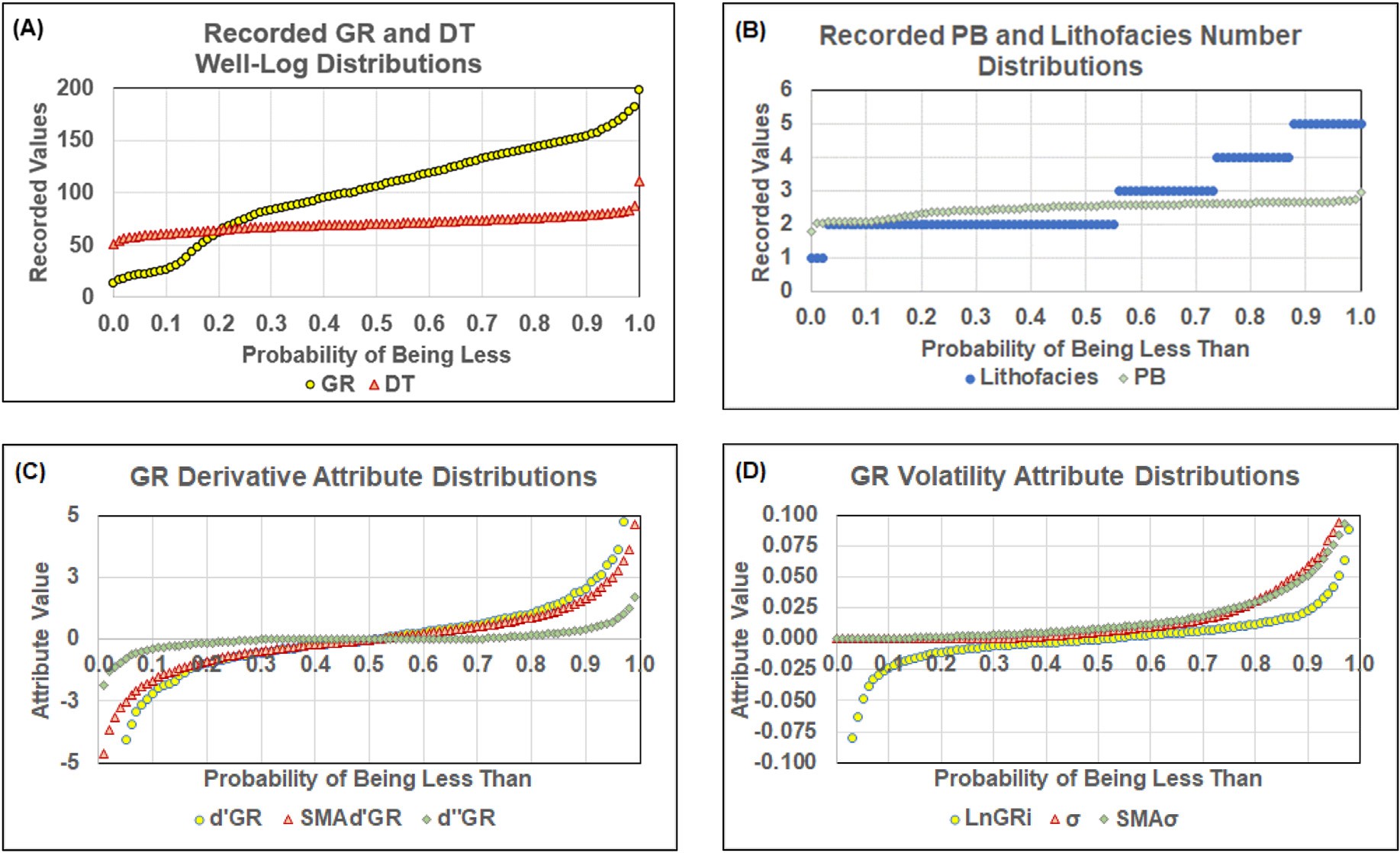


Fig. 5. Cumulative probability distributions for: (A) recorded GR and DT logs; (B) recorded PB log and lithofacies number; (C) calculated GR derivative attributes; and,

(D) calculated GR volatility attributes. The extreme ends of some of the GR derivative and GR attribute distributions are excluded to facilitate displays on mean- ingful scales.

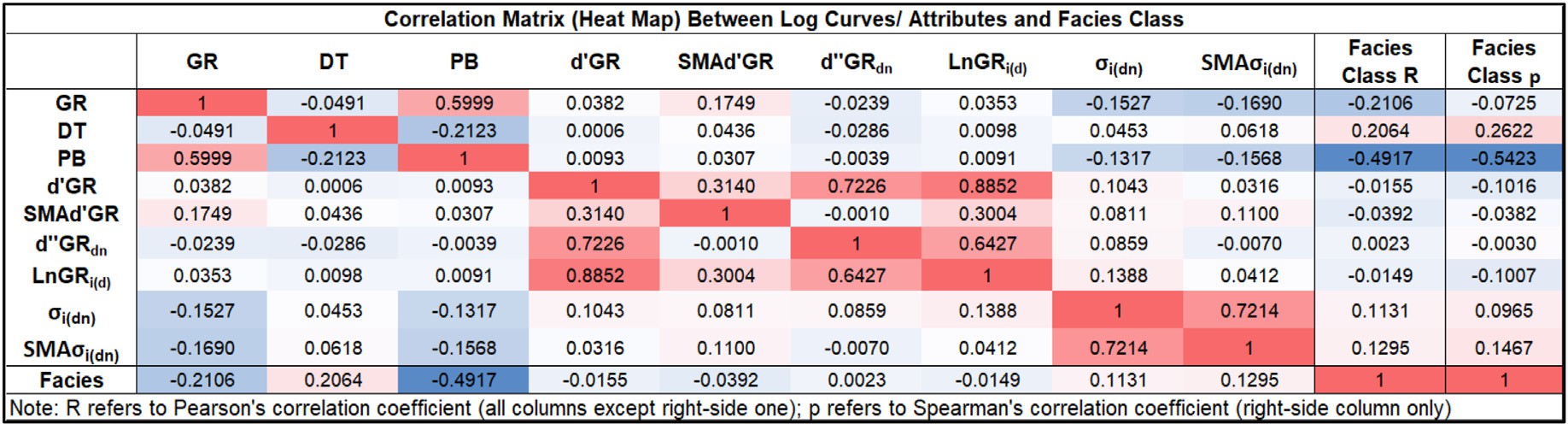


Fig. 6. Pearson's Correlation coefficients for variables associated with the Triassic well log interval of well WK-1. Both Pearson's and Spearman's coefficients are displayed for correlations of the variables with facies class.

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](#_bookmark14).

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 ach- ieved 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 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

P

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 configu- rations evaluated and compared are:

* 9-var: (*GR) (DT) (PB) (d*’*GR) (SMAd'GR) (d*’‘*GR)* (*LnGRi) (σGRi)*

*(SMAσi(dn*))

* 7-var: (*GR) (d*’*GR) (SMAd'GR) (d*’‘*GR)* (*LnGRi) (σGRi) (SMAσi(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 configu- rations 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 pre- dict 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

reproduceable 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](#_bookmark15).

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 time 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

Multi-layer Perceptron (MLP)

Ordinary Least Squares Regression (OLR)

activation fn. = relu; Solver = adam; alpha = 0.001 3 hidden layers with 100, 50 and 25 neurons; for models 2-var, 3-var, 9-var; alpha = 0.001 for

models 1-var and 7-var; adaptive learning rate

No control parameters required

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

Random Forest (RF) Number of estimators = 750; maximum depth = 150; splitting criteria = gini

cross-validation statistics can be usefully compared ([Table 4](#_bookmark15)).

[Table 4](#_bookmark15) displays the mean MAE values and MAE's standard deviations

Stochastic Gradient Descent Classifier (SGD)

Support Vector Classifier (SVC)

Learning rate/regularization α = 0.0001; maximum iterations = 10000

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

(σ2MAE) 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 σ2MAE low in all cases, and more than an order of magnitude

lower than the mean MAE values. This demonstrates that all the models

XG Boost (XGB) Number of estimators = 2000; Maximum depth = 7;

eta = 0.01; Subsample = 0.7; Columns sampled per tree = 0.5

* 1. *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/

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 σ2MAE 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 σ2MAE values for those variable configurations. Also the XGB model is not among the best performing models for the var-3 and var-2 configu- rations. 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)

Training Subset (80%) Testing Subset (20%) Applied to All 8911 Data Records

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | R2 | RMSE | MAE | ∑Error |  | R2 | RMSE | MAE | ∑Error |  | R2 | RMSE | MAE | ∑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 (σ2) 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)

9 Input Variables 7 Input Variables 3 Input Variables 2 Input Variables 1 Input Variables

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | MAE | σ2MAE |  | MAE | σ2MAE |  | MAE | σ2MAE |  | MAE | σ2MAE |  | MAE | σ2MAE |  |
| 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 σ2MAE 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 per- forms well in lithofacies classification for the var-1 (GR only) configu- ration 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 configura- tion than they do for the 3-var configuration. Nevertheless, this general ranking trend observed, based on [Table 4](#_bookmark15) 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.

* 1. *Detailed lithofacies prediction results for the testing subsets*

A clearer understanding of the relative lithofacies prediction perfor- mances 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](#_bookmark22).

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)  RF  Test R2 | 7.1  0.9604 | 6.6  0.9078 | 5.6  0.6726 | 5.7  0.4531 | 5.3  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)  SVC  Test R2 | 113.6  0.9248 | 94.3  0.6845 | 75.4  0.6701 | 61.3  0.4603 | 92.3  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)  XGB  Test R2 | 20.2  0.9609 | 52.7  0.9351 | 29.3  0.4687 | 83.1  0.4098 | 192.7  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](#_bookmark16).

Those results for the other poorer performing models are shown in [Appendix B](#_bookmark22) [Table B1](#_bookmark25). Considering the four performance metrics (R2, RMSE MAE and Error) together provides a clearer indication of how each model is performing with respect to the modelled dataset. The re- lationships between these performance measures are shown in [Figs. 7](#_bookmark17)–[9](#_bookmark17). It is apparent from [Table 4](#_bookmark15) 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 re- cords 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 Error. However, the RF model generates slightly lower RMSE and MAE values, and a slightly higher R2 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

P

P

well.

P

[Fig. 7](#_bookmark17) reveals that there is a well-defined trend of increasing Error with MAE among the different models evaluated. Excluding configura- tion var-1, the OLR and SGD model deliver poorer prediction perfor- mance than the ML models for all other variable configurations.

[Fig. 8](#_bookmark18) displays, on an expanded scale, only the high performing end of the trend shown in [Fig. 7](#_bookmark17). 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 con- figurations compared to the var-3 configuration. [Fig. 9](#_bookmark19) reveals the same result for R2 versus RMSE with the ML models following a well-defined trend.

The results presented in [Table 4](#_bookmark15) and [Figs. 7](#_bookmark17)–[9](#_bookmark17) therefore confirm the findings of the 5-fold cross validation analysis ([Table 3](#_bookmark14)). 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 pro- vides 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 wellbores 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.

1. Discussion
   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](#_bookmark22). The value that such displays provide is illustrated by [Fig. 10](#_bookmark20) 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 asso- ciated with misclassification of facies class 2 ([Fig. 10](#_bookmark20)A). 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 mis- classifications ([Figs. 10](#_bookmark20)B), 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,

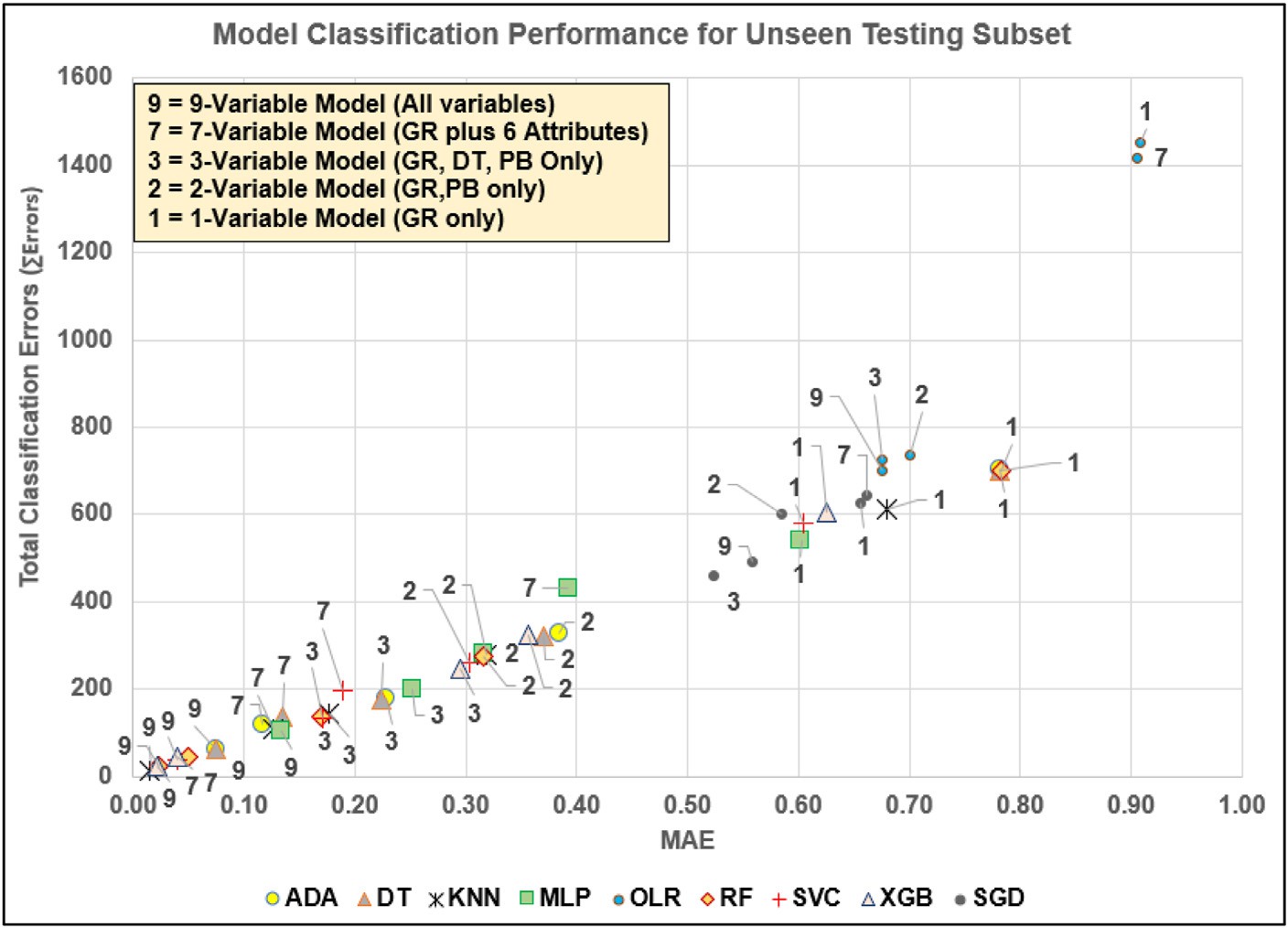


Fig. 7. PError versus MAE for the testing subsets of nine prediction models applied to the five input variable configurations evaluated, labelled 9, 7, 3, 2, and 1.

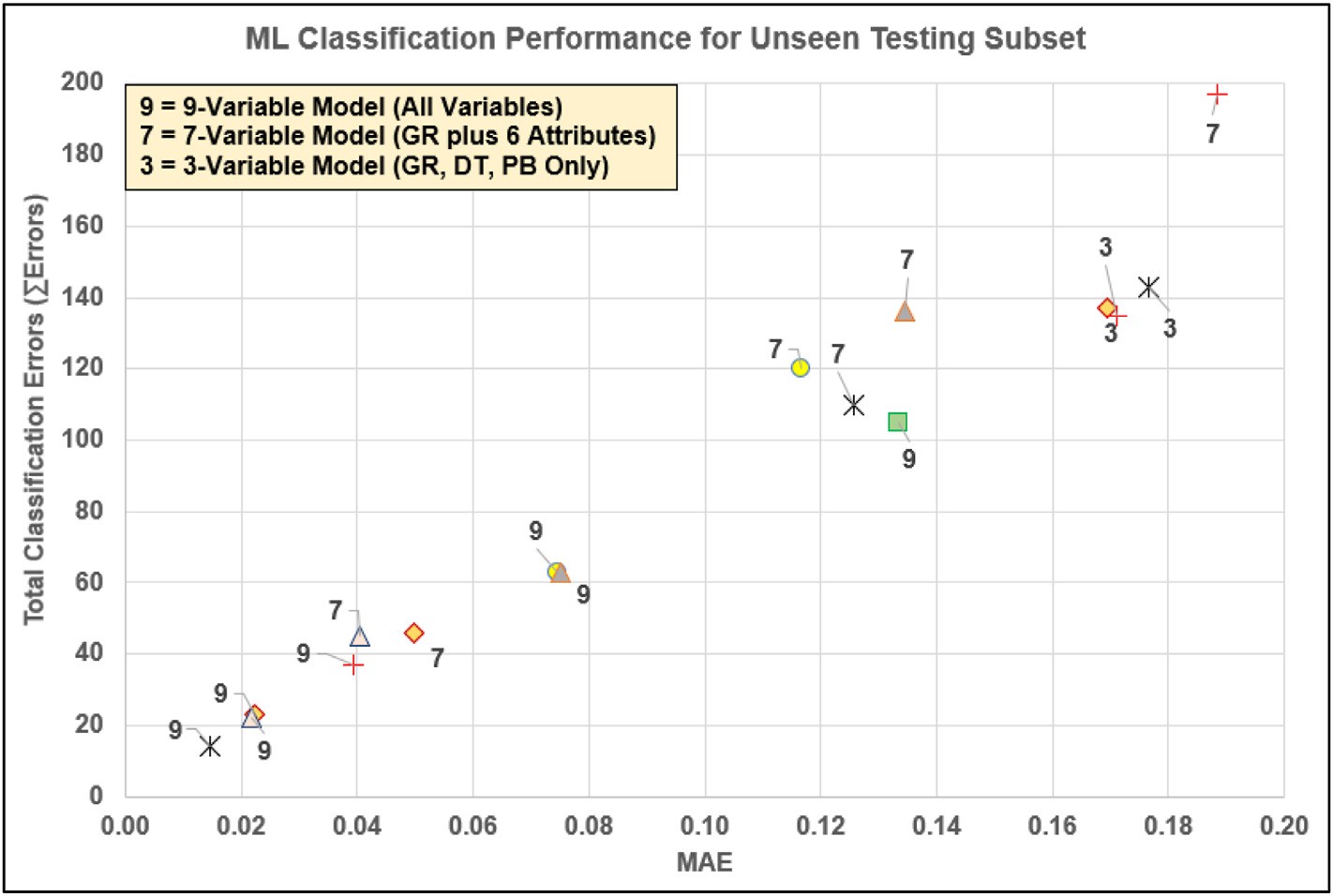


Fig. 8. PError versus MAE for the testing subsets of the best performing classification models (Var-9, Var-7 and Var-3). Symbols are the same as those used in [Fig. 6](#_bookmark12).

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 mis- classifications ([Figs. 10](#_bookmark20)C), 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 pre- diction 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](#_bookmark21) presents such displays for the most ac- curate ML models (KNN; XGB; RF), those generating the least misclas- sification, for configurations 9-var, 7-var and 3-var, respectively.

[Fig. 11](#_bookmark21)A displays the results of each data record for the KNN model

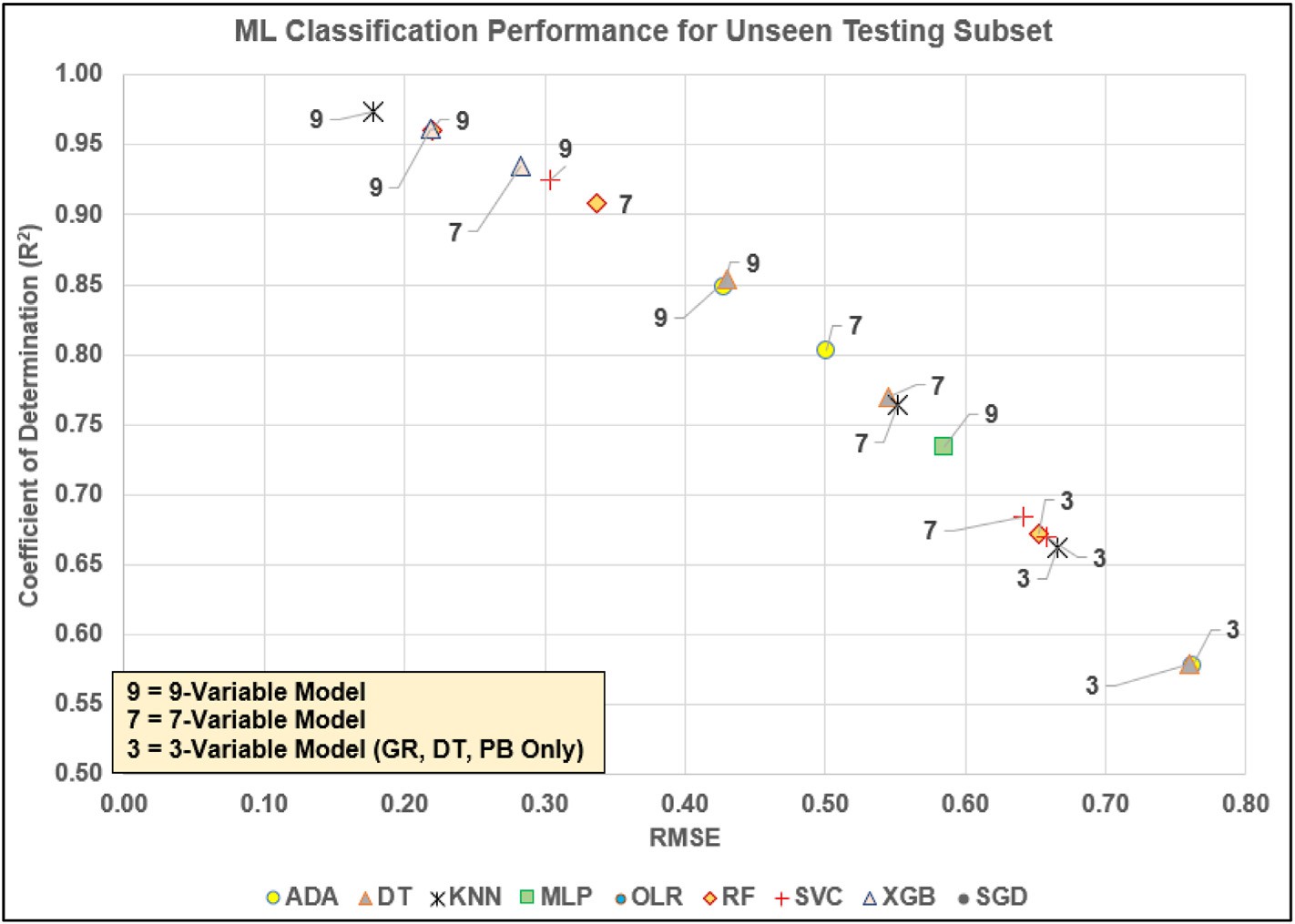


Fig. 9. R2 versus RMSE for the testing subsets for best performing classification models (Var-9, Var-7 and Var-3).

applied to the 9-var model, highlighting the impressive prediction per- formance achieved (just 14 errors among 8911 data records). [Fig. 11](#_bookmark21)B 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. 11](#_bookmark21)C, 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. 10](#_bookmark20)C.

[Figs. 10 and 11](#_bookmark20) explain why the inclusion of the GR attributes im- proves 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](#_bookmark23) 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 mis- classifications 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 tran- sition zones that are the hardest for ML algorithms to classify correctly.

* 1. *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](#_bookmark11)) 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 lith- ofacies 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 out- performed 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 vari- able when formulating its lithofacies predictions. For the dataset evalu- ated KNN's performance improves progressively as more input data variables become available ([Table 5](#_bookmark16)). 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 dis- tributions 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 configura- tion. 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](#_bookmark11)). It seems that in the absence of the PB

data the MLP and SVC models' prediction performances decline sub-

stantially. 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

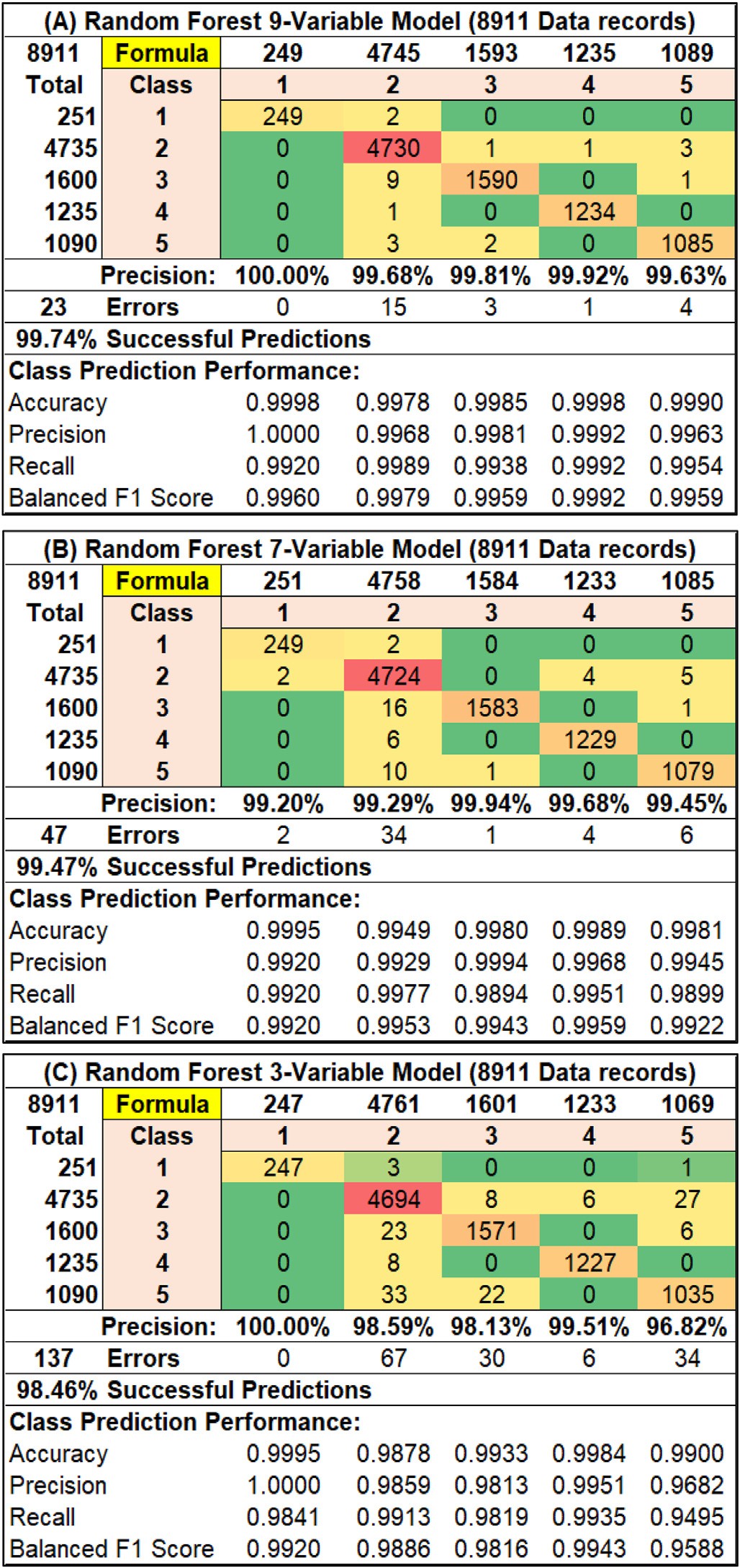


Fig. 10. Annotated confusion matrices displaying misclassification error anal- ysis for the 9-var, 7-var and 3-var models evaluated with the random forest model for all data records.

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 per- forming models, such as RF and KNN ([Table 4](#_bookmark15)). 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 data- sets, evaluating an ensemble of ML models is typically not too onerous a task in terms of computational effort ([Table 2](#_bookmark13)).

The ML models evaluated in this study all, to a degree, lack trans- parency when it comes to revealing exactly how the facies class pre- dictions 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](#_bookmark84)) has demon- strated its ability to both predict and data mine lithofacies datasets ([Wood, 2019](#_bookmark85)). 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.

1. 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 character- istics from GR curves that are not captured by the recorded log values. The first and second derivatives plus a moving average of the first de- rivative 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 rele- vant 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 ma- chine 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 misclassifi- cation from 8911 data records. This compares with 135 misclassification for the best 3-var model (Support Vector Classification (SVC)). Even more

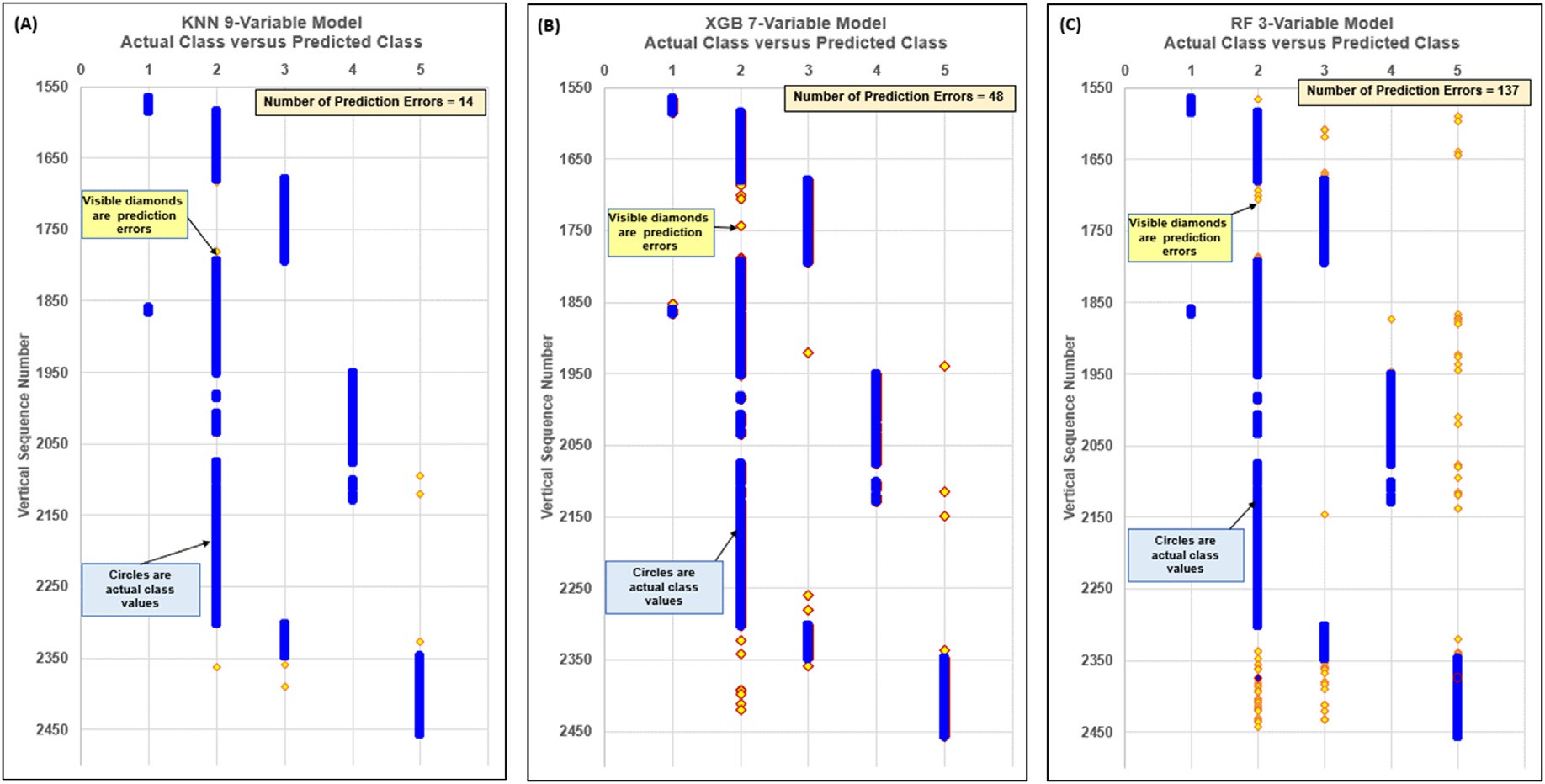


Fig. 11. Observed (blue dots) versus ML-predicted (yellow diamonds) facies lithofacies classes for all 8911 data records of WK-1 Triassic section for the best per- forming models: (A) KNN for the 9-variable model; (B) XGB for the 7-variable model; and, (C) RF for the 3-variable model. At depths where 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)](#_bookmark20) 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 configura- tion. 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 correla- tions, KNN/RF/XGB, clearly provide superior lithofacies prediction per- formance. 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 in- terests 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](#_bookmark24).

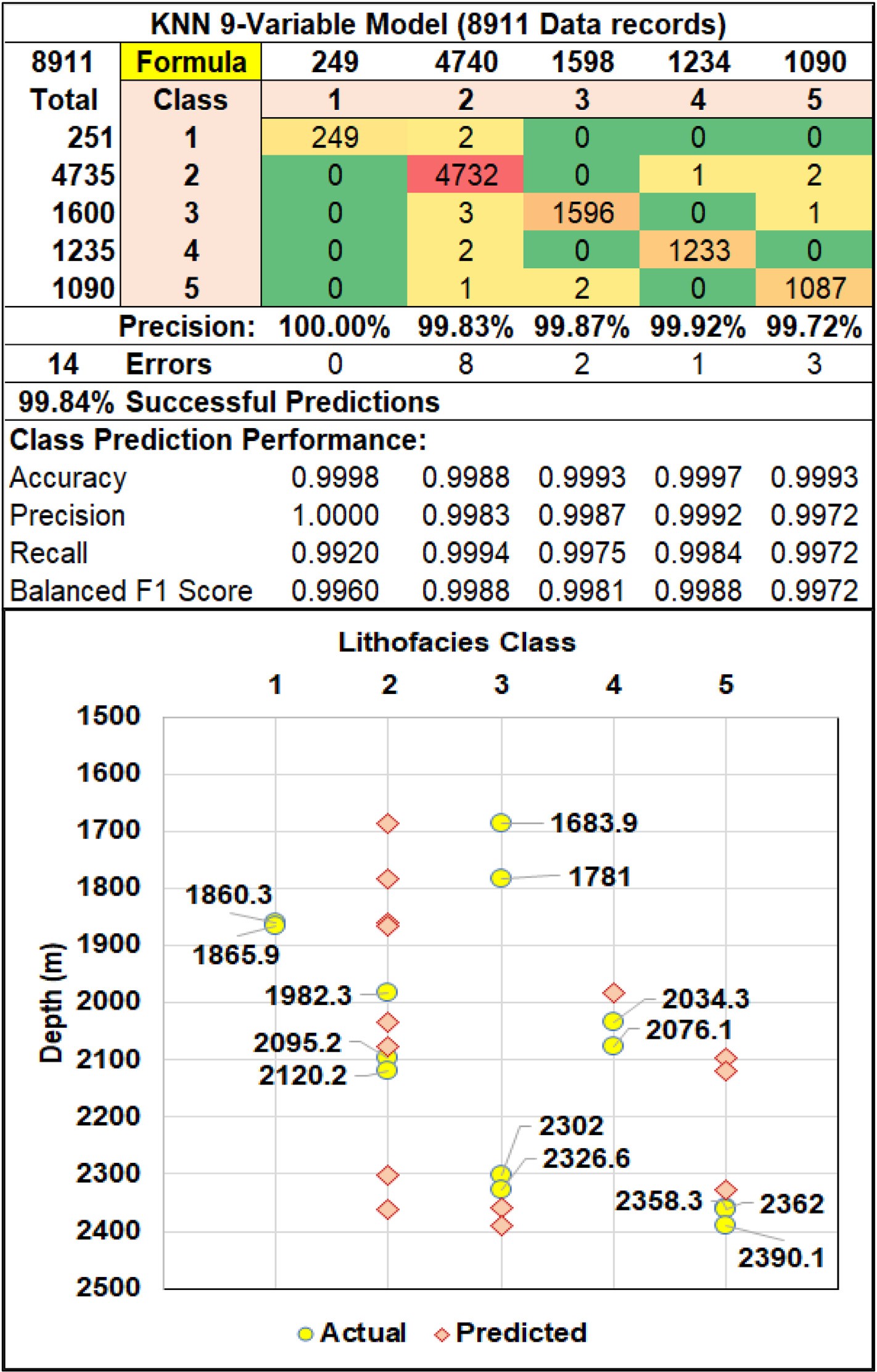


Fig. 12. Misclassification analysis for the best-performing KNN model (KNN) applied to the 9-var dataset configuration. The labels in the lower graphic represent the depth values of the fourteen misclassified data records.

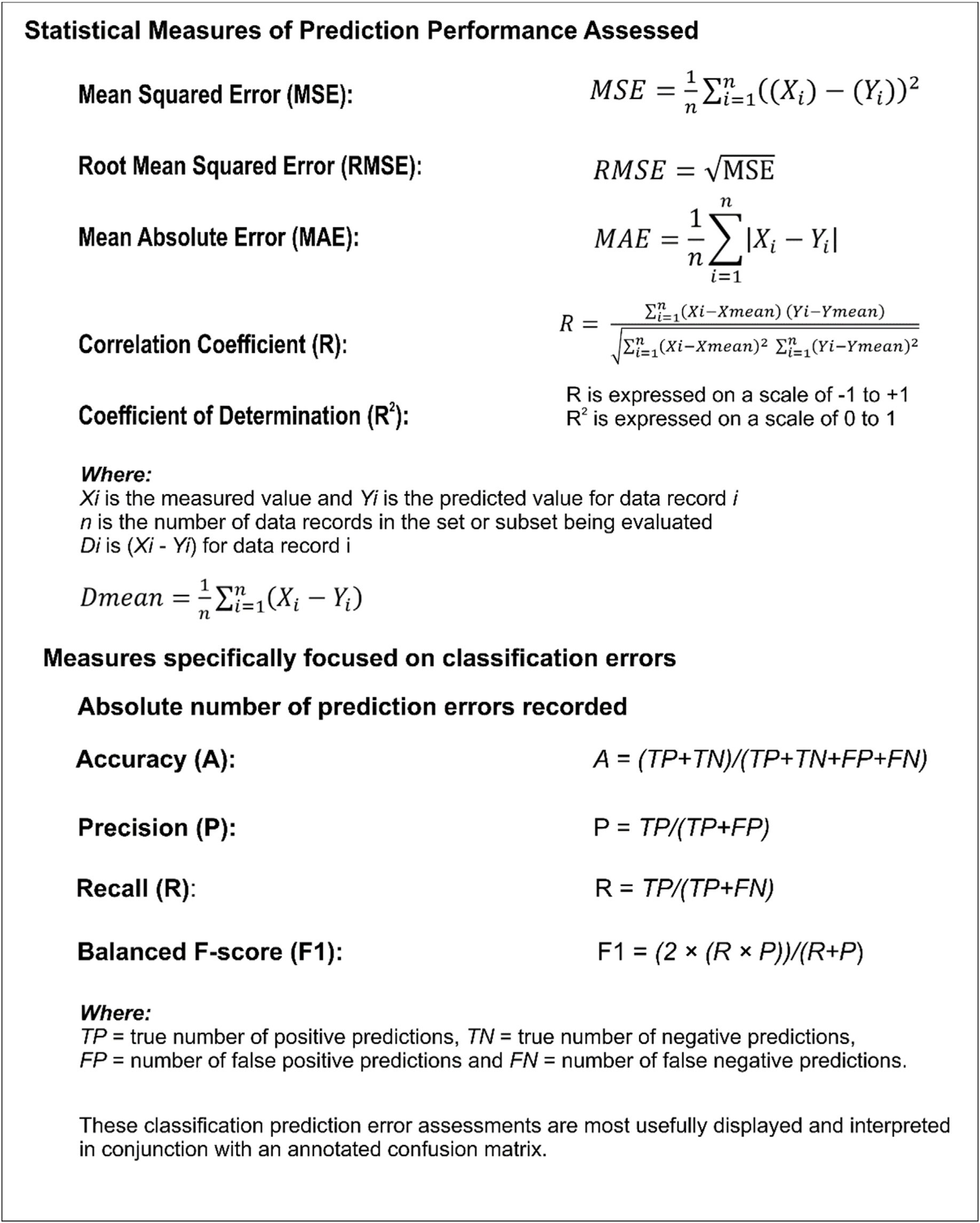


Fig. A1. Statistical measures of prediction performance appropriate for both regression and classification analysis.

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](#_bookmark25)) 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](#_bookmark16).

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)  DT  Test R2 | 5.7  0.8541 | 5.6  0.7694 | 5.5  0.5787 | 5.8  0.3931 | 199.9  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)  OLR  Test R2 | 5.6  0.2730 | 5.7  0.0523 | 5.5  0.2671 | 5.5  0.2535 | 6.1  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)  MLP  Test R2 | 4.9  0.7340 | 5.3  0.3943 | 5.2  0.5363 | 5.0  0.4521 | 4.7  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)  SGD  Test R2 | 210.6  0.1074 | 287.6  0.0623 | 208.2  0.1334 | 129.5  0.1516 | 15.2  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/](https://doi.org/10.1115/1.4051573) [10.1115/1.4051573](https://doi.org/10.1115/1.4051573), 043003.

Al-Mudhafar, 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](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref4) [Neural Networks. Cambridge University Press, ISBN 978-0-521-65263-6](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref4).

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](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref6) [Models, Response to Sea Level Changes. Geol. Assoc., Canada, pp. 27](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref6)–[45](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref6).

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](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref9) [rock facies classification problem. Comput. Geosci. 33, 599](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref9)–[617](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref9).

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/](https://doi.org/10.2118/203726-MS) [203726-MS](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/](https://doi.org/10.2118/60305-MS) [10.2118/60305-MS](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/](https://doi.org/10.1007/s40808-016-0165-z) [10.1007/s40808-016-0165-z](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.marpetgeo.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/](https://doi.org/10.1109/LGRS.2020.2968356) [10.1109/LGRS.2020.2968356](https://doi.org/10.1109/LGRS.2020.2968356).

[Fix, E., Hodges Jr., J.L., 1951. Discriminatory Analysis, Nonparametric Discrimination:](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref17) [Consistency Properties. USAF School of Aviation Medicine. Technical Report](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref17).

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/](https://doi.org/10.1006/jcss.1997.1504) [10.1006/jcss.1997.1504](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/](https://doi.org/10.1190/1.3553479) [10.1190/1.3553479](https://doi.org/10.1190/1.3553479). W1–W13.

[Goldberger, A.S., 1964. Classical Linear Regression. Econometric Theory, vol. 158. John](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref20) [Wiley](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref20) & [Sons, New York, ISBN 0-471-31101-4](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref20).

[Goncalves, C.A., Harvey, P.K., Lovell, M.A., 1995. Application of a multilayer neural](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref21) [network and statistical techniques in formation characterization. In: SPWLA 36th](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref21) [Annual Logging Symposium, Paris, 26](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref21)–[29 Jun, Society of Petrophysicists and Well](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref21) [Log Analysts, p. 12. Houston (U.S.A.)](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref21).

Hall, B., 2016. Facies classification using machine learning. Lead. Edge 35 (10), 906–909.

<https://doi.org/10.1190/tle35100906.1>, 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/](https://doi.org/10.1190/tle36030267.1) [tle36030267.1](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/](https://doi.org/10.1007/s11004-019-09838-0) [10.1007/s11004-019-09838-0](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/](https://doi.org/10.1109/34.709601)

[34.709601](https://doi.org/10.1109/34.709601).

Hossein, 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](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref27) [lakes. Chapter in. In: Brenchley, P.J., Rawson, P.F. (Eds.), The Geology of England](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref27) [and Wales. Geological Society of London](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref27).

[Howard, A.S., Warrington, G., Ambrose, K., Rees, J.G., 2008. A Formational Framework](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref28) [for the Mercia Mudstone Group (Triassic) of England and Wales. British Geological](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref28) [Survey, Research Report, RR/08/04](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref28).

[Hurst, A., 1990. Natural gamma-ray spectroscopy in hydrocarbon bearing sandstones](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref29) [from the Norwegian continental shelf. In: Hurst, A., Lovell, M.A., Morton, A.C. (Eds.),](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref29) [Geological Applications of Wireline Logs. Geological Society, London, Special](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref29) [Publication, vol. 48, pp. 211](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref29)–[222](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref29).

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](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref31) fl[uids in the Wessex basin,](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref31) [Southern England. Bull. Turk. Assoc. Petrol. Geol. 27 (1), 31](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref31)–[56](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref31).

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 classi](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref33)fi[cation using](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref33) [random forest algorithm. In: SEG International Exposition and 88th Annual Meeting,](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref33)

[pp. 2161](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref33)–[2165, 0.1190/segam2018-2998553.1](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref33).

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/](https://doi.org/10.1080/22020586.2019.12072918)

[22020586.2019.12072918](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](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref37) [lithofacies from geophysical logs: a review of methods from manual picking to](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref37) [machine learning. In: British Geological Survey (Nottingham, England) Open Rep..](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref37) [OR/21/006. 43pages](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref37).

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. Anu´ario do Instituto de Geoci^encias, 35024. [https://doi.org/10.11137/2021\_](https://doi.org/10.11137/2021_44_35024) [44\_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/](https://doi.org/10.1007/BF00116251) [10.1007/BF00116251](https://doi.org/10.1007/BF00116251).

[Reverdy, X., Argaud, M., Walgenwitz, F., 1983. Minerological Analysis Required for Log](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref41) [Interpretation in Complex Lithologies. Paper H, Transactions of the SPWLA 8th](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref41) [European Symposium](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref41).

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,](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref43) [England. Rep CF81/03. Stationery Of](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref43)fi[ce Books, p. 196, 13:978-0118841931,](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref43) [London, United Kingdom](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref43).

[Rider, M.H., 1986. Geological Interpretation of Well Logs. Blackie, ISBN 9780216918467,](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref44)

[p. 175. New York (U.S.A.)](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref44).

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/](https://doi.org/10.1144/GSL.SP.1990.048.01.04) [10.1144/GSL.SP.1990.048.01.04](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](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref46) [logs using a neural network (1). AAPG (Am. Assoc. Pet. Geol.) Bull. 76, 731](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref46)–[739](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref46).

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/](https://doi.org/10.1190/1.1445076)

[1.1445076](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.](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref49) [American Association of Petroleum Geologists Memoir, ISBN 9780891813071,](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref49)

[p. 410](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref49).

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/sklear](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html) [n.model\_selection.GridSearchCV.html](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](https://scikit-optimize.github.io/stable/modules/generated/skopt.BayesSearchCV.html)

[-optimize.github.io/stable/modules/generated/skopt.BayesSearchCV.html](https://scikit-optimize.github.io/stable/modules/generated/skopt.BayesSearchCV.html), 2nd February 2022.

SciKit Learn, 2022b. Cross-validation: Evaluating Estimator Performance. [https://scikit](https://scikit-learn.org/stable/modules/cross_validation.html)

[-learn.org/stable/modules/cross\_validation.html](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://](https://doi.org/10.1306/CE9397) [doi.org/10.1306/CE9397](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/](https://doi.org/10.2118/192930-MS) [10.2118/192930-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://](https://doi.org/10.4043/30086-MS) [doi.org/10.4043/30086-MS](https://doi.org/10.4043/30086-MS).

[Underhill, J.R., Stoneley, R., 1998. Introduction to the development, evolution and](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref58) [petroleum geology of the Wessex Basin. In: Underhill, J.R. (Ed.), Geol. Soc. Lond, vol.](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref58)

[133. Spec. Publ, pp. 1](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref58)–[18](http://refhub.elsevier.com/S2666-5441(22)00007-7/sref58).

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/](https://doi.org/10.26804/ager.2018.02.04) [ager.2018.02.04](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/](https://doi.org/10.46690/ager.2022.01.06) [ager.2022.01.06](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-](https://doi.org/10.1190/GEO2017-0595.1) [0595.1](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/](https://doi.org/10.1155/2019/5309852) [10.1155/2019/5309852](https://doi.org/10.1155/2019/5309852).