

Original article

Well-log attributes assist in the determination of reservoir formation tops in wells with sparse well-log data

David A. Wood^{®*}

DWA Energy Limited, Lincoln LN5 9JP, UK

Keywords:

Systematic well-top picking
well-log attributes
optimized feature selection
multi-k-fold validation
formation-boundary characterization
sparse well-log suites

Cited as:

Wood., D. A. Well-log attributes assist in the determination of reservoir formation tops in wells with sparse well-log data. *Advances in Geo-Energy Research*, 2023, 8(1): 45-60.
<https://doi.org/10.46690/ager.2023.04.05>

Abstract:

The manual picking of reservoir formation boundaries using limited available well-log data in multiple wells across gas and oil reservoirs tends to be subjective and unreliable. The reasons for this are typically caused by the combined effects of spatial boundary complexity and limited well-log data availability. Formation boundary characterization and classification can be improved when treated as a binary classification task based on two or three recorded well logs assisted by their calculated derivative and volatility attributes assessed by machine learning. Two example wellbores penetrating a complex reservoir boundary, one with gamma-ray, compressional-sonic, and bulk-density logs recorded, the other with just gamma-ray and bulk-density logs recorded, are used to illustrate a more rigorous proposed methodology. By combining attribute calculation, optimized feature selection, multi-k-fold cross validation, confusion matrices, feature-influence analysis, and machine learning models it is possible to improve the classification of the formation boundary. With just gamma-ray and bulk-density recorded well logs plus selected attributes. K-nearest neighbour, support vector classification, and extreme gradient boosting machine learning models are able to achieve high binary classification accuracy: greater than 0.97 for training/validation in one well; and greater than 0.94 for testing in another well. extreme gradient boosting feature-influence analysis reveals the attributes that are the most important in the formation boundary predictions but these are likely to vary from reservoir to reservoir. The results of the study suggest that well-log attribute analysis, combined with machine learning has the potential to provide a more systematic formation boundary definition than relying only on a few recorded well-log curves.

1. Introduction

Over the past several decades, quantitative reservoir formation evaluation based on the interpretation of well-log data has focused on providing a detailed analysis of the key reservoir properties, including porosity, fluid saturations, and permeability (Serra, 1986; Darling, 2005). Such information has provided vital inputs to reservoir models and simulations (Luthi, 2001). Information from more detailed logging tools can also be used to provide analysis of a formations pore-size distributions (Liu and Ostadhassan, 2019) and a reservoir's flow unit characterization (Rafiei and Motie, 2019) and sub-facies classification (Zhang and Zhang, 2021). However, alongside such formation-character analysis, well-log data is routinely used for well-to-well correlations and the precise

picking of key formation tops calibrated with core data (both whole cores and side-wall cores). Formation-top picking based on well logs is traditionally conducted manually and is fraught with subjectivity and bias, as well as being time-consuming when tens or hundreds of wells are involved. Many efforts have been made to make this latter application more efficient with the aid of knowledge-based systems and techniques (Lineman et al, 1987). The dynamic depth-warping technique is widely used to assist well-to-well correlations and is also applied to calibrate well-to-seismic correlations (Luo et al., 2018). However, such efforts tend to be hampered by two key limitations: the limited suites of recorded well logs in most field development wells; and the transitional nature of formation boundaries between cap rock and reservoir in a

substantial number of gas and oil fields.

There is substantial interest in establishing methods that can provide automated and objective formation top correlations incorporating multiple wells (Hong and Kang, 2020) up to many hundreds of wells (Zoraster et al., 2004) or more directly while drilling (Al-AbdulJabbar et al., 2018). This automated requirement has become even more pertinent since the extensive development of unconventional gas and oil resources on a basin-wide scale involving thousands of wells tied to seismic data (Grant et al., 2018). As well as a requirement for field development studies, automated well-picking is also a potential benefit for basin-wide studies. Recently, Zhang et al. (2022) applied a dynamic time warping technique to auto-pick the tops of a specific formation of interest in about 71,000 wellbores (about 40% with core information) distributed across the Western Canada Sedimentary Basin. They did this using just the gamma ray (GR) log curve. The limited information available from a single well log curve, clearly, also represents a constraint to the accuracy achievable by such automated-picking methods.

In gas and oil field developments, for cost reasons, cores and comprehensive logging suites tend to be recorded in only a few selected wellbores. For instance, a conventional reservoir constituting a large field may involve from fifty or more (offshore) to many hundreds wellbores (offshore). Unfortunately, substantially less than ten percent of those wells are typically sampled by comprehensive suites of well logs, specialist well logs, and/or cores. The remainder of the well bores are typically sampled by just a basic suite of well logs, e.g., GR, plus compressional sonic travel time (DT), bulk density (PB), deep resistivity and/or caliper logs. For complex reservoir-cap-rock formation boundaries it is often difficult to confidently define the depth to the top of the reservoir and characterize the reservoir formation based on only a basic suite of well logs.

A recently developed technique that extracts various derivative and volatility attribute information from the GR curve (Wood, 2022a) for exploitation by machine learning models to predict lithofacies, offers the potential to be adapted for the systematic picking of specific formation tops across an area or field of interest. Indeed, this technique has been successfully extended to extract attributes from other recorded well logs to improve lithofacies prediction accuracy in various complex lithology sequences (Wood, 2022b, 2022c). This study further develops the well-log-attribute method, with the involvement of optimizers, into a technique specifically to assist in the picking of formation tops and improve formation boundary characterization in situations where limited well log and core data is available. The task is configured as a binary classification problem and the efficiency and accuracy of the technique are improved by incorporating: (1) optimizers to assist machine learning (ML) algorithms with log-feature selection; (2) multi-cross-fold validation to determine the best data splits for ML model training and validation; (3) confusion matrices to determine in which formation prediction errors are concentrated; and (4) feature influence analysis combined with correlation analysis to establish the most influential well-log features in discriminating between the two formations

involved. The technique is illustrated and tested using three and two recorded well-log curves from two wells penetrating a complex transitional cap-rock/reservoir formation boundary. Well-log attributes as described and implemented in this study are shown to add a useful dimension with which to improve the accuracy of formation-boundary definitions and characterizations.

2. Method

The challenge addressed in this study is to improve the delineation of complex formation boundaries encountered in wellbores, particularly those between cap rocks and reservoir formations, based on a limited suite of recorded well logs. This is a common task required in field development operations, where, for large gas and oil fields (conventional and unconventional) many tens of wellbores penetrate this key formation boundary, but for most of them no cores are recovered and a very sparse suite of well logs are recorded. The industry is striving to find ways to automate the picking of reservoir tops on a field wide basis but, as historical studies have demonstrated (Zhang et al., 2022), this is very difficult to achieve based on limited recorded well-log data.

This study builds on the recently developed method of calculating several derivative and volatility attributes from selected recorded well logs to improve lithofacies characterization (Wood, 2022a). This method has now been successfully applied to delineate and characterize a range of complex lithological sequences in various clastic and carbonate sedimentary sequences (Wood, 2022b, 2022c, 2022d). This method can be adapted to provide more detailed evaluation of a single boundary between two key formations (cap-rock and reservoir) rather than to distinguish multiple lithologies penetrated over more extensive depth intervals.

2.1 Workflow requirements

The technique is designed to operate generically and can be adapted to work with different types of complex formation boundaries. As such, it offers the potential to automate systematic formation boundary picking and characterization across field areas incorporating multiple wells in limited recorded well log suites. The technique proposed to implement such delineations involves a workflow sequence of steps described here and summarized in Fig. 1.

Step 1: Select the logged depth interval.

The selected logged depth interval that adequately straddles the formation boundary of interest providing about 500 to 1,000 recorded well-log depth points above and below that formation boundary. This is required for one or, ideally, more wells with core analysis information available to precisely define the formation contact depth. These wells will then form the training and validation dataset with which to calibrate the model.

Step 2: Select available recorded well logs.

Select at least two or three available recorded well logs displaying some contrast between the two formations (e.g., cap rock and reservoir) making them suitable for well-log attribute calculations. Calculate six, or more, derivative and

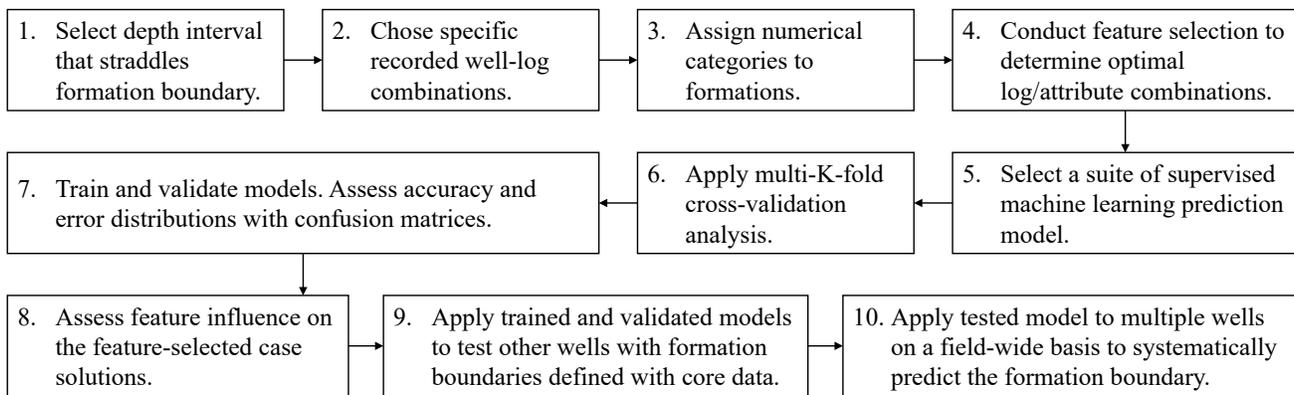


Fig. 1. Workflow diagram for a technique involving well-log attributes to pick and characterize formation boundaries.

volatility based attributes for the selected well logs, using the formulas listed in Supplementary File: Section S7, to extract more textural information from each recorded well log. GR, DT and PB are often suitable candidates but other recorded logs available in most of the wells drilled throughout an oil-/gas-bearing reservoir may be worthwhile considering for this purpose. Whatever well-logs are selected it is necessary to make sure that they have been adequately corrected for borehole-environment conditions to avoid anomalous influences from drilling/borehole conditions. There are various statistical methods used to identify and repair/filter poor well-log data (Banas et al., 2021).

Step 3: Assign numerical formation category identifiers.

The numerical formation category identifiers are assigned to the upper formation (e.g., cap rock) and lower formation (e.g., reservoir). Useful category numbers to use for this purpose are (-1) for the formation above the contact boundary and $(+1)$ for the formation below the formation boundary, making it a distinct binary classification task. This numerical categorization is useful for two reasons:

- 1) It facilitates the calculation of Pearson's correlation coefficients between recorded the well logs, their calculated attributes and the two formation categories.
- 2) It can be conveniently exploited to generate consistent numerical error measurements when predictions are compared with actual category values. For instance, subtracting correct predictions from actual values generates an error value of zero, whereas an incorrect absolute error generates a values of two, and a squared error value of four.

Step 4: Conduct feature selection.

This step identifies combinations of recorded well logs and certain well-log attributes that generate high accuracy when predicting the numerical formation categories. The K-nearest neighbour (KNN) machine-learning algorithm combined with a number of optimizers is a useful approach to adopt for this purpose (Wood, 2022d), because it is fast to implement. The benefit of using multiple optimizers is that each optimizer is based on a different algorithm enabling them collectively to explore the feasible solution space more efficiently and potentially generate more accurate predictions. This step also

identifies the most effective well log attributes in distinguishing the formations. Six optimizers were used in this study to conduct feature selection. These are: differential evolution (DE) (Zhang et al., 2020); cuckoo search optimizer (CSO) (Wood, 2016); Jaya (Rao, 2016), particle swarm optimizer (PSO, Atashnezhad et al., 2014); sin-cosine algorithm (SCA) (Abualigah and Diabat, 2021); and Salp swarm algorithm (Faris et al., 2020). These optimizers were selected based on their rapid execution times and accuracies observed with a range of lithological datasets. The KNN-optimizer models are applied with Python code to minimize a cost function that introduces an incentive (Supplementary File: Section S8) to favor solutions with the smallest number of features, while maintaining prediction accuracy (Wood, 2022d).

Step 5: Select formation prediction models to evaluate and compare.

It is beneficial to apply several supervised ML models and compare their prediction performance with that of a classification regression model (e.g., logistic regression) (Cox, 1958). As well-log variables and lithology tend to be related with a degree of non-linearity, ML algorithms tend to be able to make more accurate formation prediction than logistic regression, which does assume linearity between the independent variables (i.e., well-log distributions in this task) and the probability distributions used to make its discrete binary classifications. For this study, KNN (Fix and Hodges, 1951), support vector classification (SVC) (Cortes and Vapnik, 1995) and extreme gradient boosting (XGB, Chen and Guestrin, 2016). These three ML models are selected because they exploit distinct mathematical methodologies, execute rapidly with datasets of moderate sizes, and, based on lithofacies studies conducted with many datasets. The prediction-classification algorithms were executed using Python code incorporating customized SciKit Learn functions (SciKit Learn, 2023a).

Each ML prediction-classification model involves model-specific control parameters that require tuning for each dataset involving different well-log feature combinations, Grid-search analysis (SciKit Learn, 2023b) and Bayesian optimization (SciKit Learn, 2023c) were conducted in conjunction with trial-and-error tests to determine the optimum values for the KNN, SVC and XGB models.

- 1) KNN applies data-record matching to determine a specified number of the closest matching data records in a model training dataset, i.e., the “K” nearest neighbours). The key control variables to be tuned are the value of K and the Minkowski-distance formula to be applied (typically that involves a choice between Euclidian or Manhattan distance options; Shahid et al., 2009). For the feature-selected datasets assessed, the optimum K value varied from 5 to 10 and the Euclidian distance measure provided the most accurate predictions in all cases.
- 2) SVC determines the optimum support vectors required to transform the dataset variables into into multi-dimensional hyperspace. The selection of the most effective kernel function is an SVC tuning requirement. Trial-and-error tests revealed that the radial basis function (RBF) kernel worked best with all the datasets evaluated in this study, which is typically the case for datasets that involve non-linearity (Chang et al., 2010). Other SVC control parameters to be tuned include an error-regularization factor (C), the RBF’s depth-of-influence factor (γ), and an error-tolerance limit (ϵ) above which cost-function penalties are applied. For the datasets evaluated C varied between 700 and 750, the default γ option “scale” and the ϵ value 0.001 where applied in all cases.
- 3) XGB applies gradient boosting to optimize the prediction-classification performance of an ensemble of decision trees (Chen and Guestrin, 2016). Successive iterations work to improve upon the classification accuracy achieved collective ensemble of decision trees. The algorithm involves a number of control parameters, which make it flexible but somewhat time consuming to optimally tune. Key tuneable parameters (with the range of optimal values used in this study shown in brackets) are: number of decision trees (750 to 1,000), maximum tree depth (15 to 20), learning rate η (0.01 to 0.1), subsample limit of data records sampling for each decision tree (0.7 to 0.9), and the fraction of features (columns) to be considered by each tree (0.8 to 1.0).

The binary selection made by the logistic regression algorithm worked best applying an ElasticNet regularization function with an L1 ratio of 0.5 and a regularization tolerance limit of 0.0001.

Step 6: Conduct multi-k-fold cross-validation analysis.

Each prediction model is analysed with multi-k-fold cross-validation for a range of cases compiled with different feature-optimized combinations. Some of the cases selected should involve features related to just pairs of recorded well logs, that have been recorded in multiple wells throughout the reservoir. This enables the relative prediction performance of pairs of well logs and their attributes to be assessed. By comparing 4-, 5-, 10- and 15-fold evaluations, each repeated with several runs, the most effective data splits to use for the training and validation subsets can be determined. The optimum splits may vary depending upon the feature combinations selected for each case. Useful information can be obtained by assessing the multi-k-fold cross-validation results in terms of the mean and standard deviations accuracy and error metric

values determined from the results of multiple runs to make those values statistically meaningful. The fold that generates the highest mean minus standard deviation accuracy value typically indicates the best performing fold for a dataset. Hence, if the 10-fold analysis provides the highest mean minus standard deviation accuracy value that suggests that random dataset splits of 0.9 to the training subset and 0.1 to the validation subset should generate the most reliable prediction results.

Step 7: Train and validate random data-record subsets.

Each feature-selected dataset case is used to train and validate using random data-record subsets with each of the selected formation-classification prediction models, incorporating data from one or more wells in which the formation boundary is reliably established (e.g., from core analysis or spectral gamma-ray well logs). Compare the prediction results using accuracy as the primary prediction performance criteria, verified with various error metrics. Confusion matrices should also be constructed to establish the distribution of prediction errors between the two formations considered. Depth versus error plots that identify the exact depth positions of erroneous formation predictions within the two formations reveal the problematic zones, in one of both of the formations, that are most difficult for the prediction models to distinguish.

Step 8: Conduct feature importance analysis.

The feature importance analysis aims to determine the most influential recorded well logs and attributes associated with trained model solutions for multiple feature-selected cases. XGB and logistic regression readily reveal such information in fractional terms. Compare the results with the relative distribution of Pearson’s correlation-coefficient values between the selected features and the formation categories. Such analysis is valuable for identifying the selected features that contribute the most to the prediction accuracy of each case based on different feature combinations. This information provides insight to the well-log textural features that are most distinctive between the two formations involved.

Step 9: Apply trained models to data from other wells for independent model testing.

Select the best performing trained and validated prediction models and apply them to data from other wells, some with only limited well log suites recorded, but with sufficient core data, or specialized well-log data to verify the depth location of the formation boundary. Use the trained models to predict the formation categories and automatically select and characterize the formation boundary in the test well(s). Assess the prediction performance of the models applied to the test-well data in terms of accuracy and error metrics and apply confusion matrices and error versus depth plots to clarify the prediction error distributions.

Step 10: Apply trained, validated and tested model to multiple wells without core information.

Such applications of the tested model could be across an entire gas/oil field reservoir area, involving all logged wells, including those wells with limited recorded suites of well-log data.

2.2 Pre-processing of well-log datasets

The suite of well-log attributes calculated requires data from the immediately adjacent shallower depths, up to about twenty data records, in order to adequately calculate the attributes involving moving averages. Consequently, the first few data records (nineteen in this study) of the depth interval selected are omitted from datasets used in the prediction analysis.

To avoid range biases in the independent-variable feature distributions all recorded well logs and attributes are normalized, prior to input to the prediction models, to scale ranges of -1 to $+1$ by applying Eq. (1).

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

where Nx_i^m refers to the m^{th} well-log feature's normalized value. This value is derived by adjusting the features actual value (x_i^m) for the i^{th} data record by minimum (x_{\min}^m) and maximum (x_{\max}^m) values of the m^{th} features distribution.

2.3 Classification prediction performance measures

The primary prediction performance measures used in this binary classification task are accuracy and percentage error (%Error) are as follows:

$$\text{Accuracy} = 1 - \frac{n_e}{t_p} \quad (2)$$

$$\% \text{Error} = 100 \times \frac{n_e}{t_p} \quad (3)$$

where n_e is number of errors and t_p is total data records predicted.

However, by assigning the two formation categories the numerical values of -1 and $+1$, it is also possible to calculate other numerical statistical error metrics, including mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) that provide complementary prediction performance information. See Supplementary File: Section S2 for formulas used to calculate these metric.

3. Materials

Logged sections from two wellbores (Well A and Well B) penetrating the cap rock and limestone reservoir of a large oil field involving many development wells are used to illustrate the formation boundary determination method described. The cap rock is a laminated shale sequence and the reservoir, which includes highly porous limestone interspersed with shale layers of variable thickness in its upper section, The formation boundary between the cap rock and reservoir is transitional making it difficult to pick precisely across the field. The well logs available in these wells were recorded with samplings intervals of 0.1524 m (0.5 ft), resulting in about 6 data records per meter.

The Well A section selected for evaluated to train and validate the formation-boundary prediction models extends from 1,500 to 1,800 m depth with the formation boundary defined by core data located at 1640 m. The selected section is

comprised of 1,969 data records of three recorded and environmentally corrected well logs: GR, DT and PB (Supplementary File: Fig. S1). Once the well-log attributes for each of these well log are calculated, 1,950 data records are then available for the training-validation dataset (898 data records in the cap rock formation and 1,052 data records in the reservoir formation).

The Well B section selected, located 8 km from Well A, is evaluated to independently test some of the the formation-boundary prediction models trained with the Well A dataset. The Well B selected section extends from 1,600 to 1,800 m depth with the formation boundary defined by core data located at 1,680 m. The selected section is comprised of 1,314 data records of two recorded and environmentally corrected well logs: GR and PB (Supplementary File : Fig. S1). Once the well-log attributes for each of these well log are calculated, 1,295 data records are then available for the independent testing dataset (505 data records in the cap rock formation and 790 data records in the reservoir formation). The Pearson correlation coefficients for the recorded well logs and their attributes (calculated using the formulas from Supplementary File: Section S7) versus the formation category are displayed in Table 1 for Well A and Well B.

In Well A, the GR and DT log data display strong negative correlation coefficients with formation category, whereas the PB log displays a strong positive correlation coefficients (Table 1). The correlation for GR and PB with formation category are of the same type but stronger in Well B. Those correlations are easy to discern from the recorded log versus depth trends in those two wells (Figs. 2 and 3). In both wells, most of the calculated well-log attributes display poor correlations ($\ll 0.1$) with formation category. However, attributes DT5, DT6, PB5 and PB6 (related to volatility and moving-average volatility) in Well A display moderate negative correlation coefficients with formation category. In Well B, attributes PB5 and PB6 display moderate negative correlation coefficients with formation category, whereas attributes GR5 and GR6 display moderate positive correlation coefficients with formation category. These relationships suggest that the prediction models are more likely to be influenced by the volatility and moving-average volatility attributes than the other attributes.

4. Results

4.1 Prediction performance using only the recorded well logs

The feature cases evaluated by the four prediction models (KNN, LogReg, SVC and XGB) considering only the three recorded well logs are:

- 1) Case 0 (three features): GR0, DT0, PB0
- 2) Case 1 (two features): GR0, DT0
- 3) Case 2 (two features): GR0, PB0

Table 2 presents the binary formation category prediction results for a randomly selected validation subset for these three cases for the KNN, SVC and XGB models applied to the Well A data using two data splits (0.75 training: 0.25 validation; and, 0.9333 training: 0.0667 validation). The results achieved

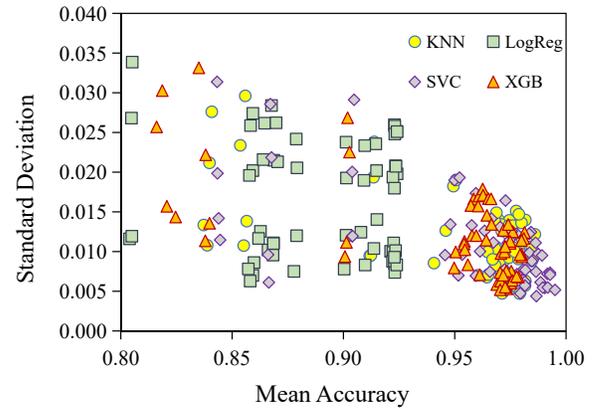
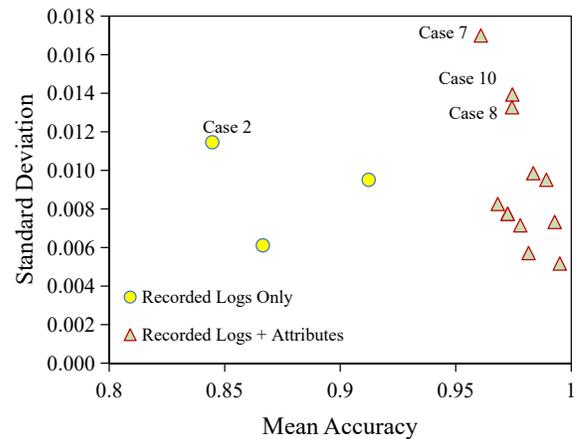
Table 1. Correlation coefficients between well logs and attributes and formation categories.

Type	Feature	Fomation category of Well A	Fomation category of Well B
GR	GR0	-0.5484	-0.7330
	GR1	0.0102	0.0073
	GR2	0.0152	-0.0134
	GR3	0.0025	0.0079
	GR4	0.0118	0.0050
	GR5	0.0158	0.3819
DT	DT0	-0.7222	/
	DT1	0.0086	/
	DT2	0.0135	/
	DT3	0.0021	/
	DT4	0.0091	/
	DT5	-0.3745	/
PB	DT6	-0.4052	/
	PB0	0.5677	0.8139
	PB1	-0.0106	-0.0109
	PB2	-0.0182	-0.0093
	PB3	-0.0030	-0.0017
	PB4	-0.0108	-0.0106
	PB5	-0.4175	-0.4542
	PB6	-0.4484	-0.4863

Notes: Well A has 1,969 data records, Well B has 1,314 data records.

by the LogReg model are not shown as they generated inferior prediction accuracies (0.75 to 0.83) to the other three models.

For Case 0 (GR0, DT0, PB0) the KNN model achieved accuracy of ~ 0.92 to 0.93 , with the SVC and XGB recording only slightly lower accuracies. For Case 1 (GR0, DT0) the SVC model achieved accuracy of ~ 0.86 to 0.87 , with the SVC and XGB recording only slightly lower accuracies. For Case 1 (GR0, PB0) the XGB model achieved accuracy of 0.82 to 0.85 , with the SVC and XGB recording similar levels of accuracy. As to be expected the prediction models achieve higher accuracy when presented with data from three logs rather than two logs. The higher correlation coefficient between DT0 and formation category compared with PB0 and formation category is consistent with the slightly higher prediction accuracies achieved by the models for Case 1 compared to Case 2. For all of these three cases the data record split of 0.75 training: 0.25 validation (488 validation subset data records) generates slightly higher prediction accuracy than the than the data record split of 0.933 training : 0.0667 validation (131 validation subset data records). However, the latter data split still provides credible and consistent levels of

**Fig. 2.** 240 cross-validation results (60 for each prediction algorithm) combining 4-fold, 5-fold, 10-fold and 15-fold results for the Well A dataset.**Fig. 3.** This graph only displays the results for the most accurate cross-validation models for each of the 15 well-log and attribute combinations (Cases 0 to 14) cases.

prediction accuracy.

The results for Cases 0, 1 and 2 (Table 2) serve as a benchmark with which to assess the performance of the cases incorporating selected well-log attributes. Although, the prediction accuracies are relatively high for Cases 0 to 2, a substantial number of prediction errors are generated for both formations (-1 and $+1$). The Cases 0, 1 and 2 prediction performances also justify the application of the KNN, SVC and XGB models for the detailed assessment of the cases including selected well log attributes. However, based on the poorer prediction performances of the LogReg models, presumably due to the linear assumptions involved, results for that model are not considered in detail for the other cases.

4.2 Feature selection applying KNN-optimizer routine

Case 3 involves all 21 features and is used as the input data set for feature selections involving GR, DT and PB logs and their attributes. The three best feature-optimized cases are

Table 2. Formation classification prediction performances achieved by ML models based on two or three recorded well logs for Well A.

Case	Model	Random split training: Validation	MAE	RMSE	R ²	Accuracy (0 to 1)	Fm (-1) correct	Fm (-1) wrong	Fm (+1) correct	Fm (+1) wrong	Total records assessed	Total errors	Error %
Case 0 GR DT PB	KNN	0.75 : 0.25	0.14754	0.5432	0.7027	0.9262	196	27	256	9	488	36	7.377
		0.9333 : 0.0667	0.1679	0.5795	0.6514	0.9160	46	7	74	4	131	11	8.3969
	SVC	0.75 : 0.25	0.1803	0.6005	0.6367	0.9098	188	35	256	9	488	44	9.0164
		0.9333 : 0.0667	0.2137	0.6538	0.5564	0.8931	45	8	72	6	131	14	10.687
Case 1 GR DT	XGB	0.75 : 0.25	0.1598	0.5654	0.6779	0.9201	197	26	252	13	488	39	7.9918
		0.9333 : 0.0667	0.1832	0.6053	0.6197	0.9084	46	7	73	5	131	12	9.1603
	KNN	0.75 : 0.25	0.2869	0.7575	0.4219	0.8566	178	45	240	25	488	70	14.344
		0.9333 : 0.0667	0.2901	0.7617	0.3979	0.8550	43	10	69	9	131	19	14.504
SVC	0.75 : 0.25	0.2623	0.7243	0.4715	0.8689	170	53	254	11	488	64	13.115	
	0.9333 : 0.0667	0.2901	0.7617	0.3979	0.8550	42	11	70	8	131	19	14.504	
Case 2 GR PB	XGB	0.75 : 0.25	0.3484	0.8347	0.2981	0.8258	178	45	225	40	488	85	17.418
		0.9333 : 0.0667	0.4122	0.908	0.1444	0.7939	45	8	59	19	131	27	20.611
	KNN	0.75 : 0.25	0.3115	0.7893	0.3724	0.8443	167	56	245	20	488	76	15.574
		0.9333 : 0.0667	0.3359	0.8196	0.3029	0.8321	40	13	69	9	131	22	16.794
SVC	0.75 : 0.25	0.3033	0.7788	0.3889	0.8484	164	59	250	15	488	74	15.164	
	0.9333 : 0.0667	0.3053	0.7815	0.3662	0.8473	42	11	69	9	131	20	15.267	
XGB	0.75 : 0.25	0.2951	0.7682	0.4054	0.8525	178	48	241	24	488	72	14.754	
	0.9333 : 0.0667	0.3664	0.8561	0.2395	0.8168	43	10	64	14	131	24	18.321	

selected to form Cases 4, 5 and 6. Seven KNN-optimizer runs (labelled A to G) were executed for each of six optimizers applied using the Well A Case 3 dataset. This amounted to forty-two optimized feature selections, which were then ranked considering both accuracy and cost function scores achieved. Table 3 displays the top eight feature selections based on their performance ranking. Each feature combination is identified by the optimizer abbreviation and its run. Hence, SCA-G (refers to seventh run of the sin-cosine algorithm). The top-three rank feature selections in Table 3 were selected for Case 4 (SCA-G), Case 5 (DE-D) and Case 6 (Jaya-C). Note the optimizer has reduced the twenty-one available features to between eight and thirteen for these three cases.

Those top-three selected cases all include GR0 and DT0 as part of their feature selection. However, only Case 5 (DE-D) includes PB0. The volatility (GR5, DT5, PB5) and rolling-average volatility (GR6, DT6 and PB6) are key selected features in most of the best performing cases (Table 3), along with the rolling average first differential (GR2, PB2). On the other hand, features GR1, GR3, GR4, DT1, DT3, DT4, PB3 and PB4 are rarely selected by the best performing feature selections derived from the Case 3 dataset.

Case 7 involves all 14 features relating to the GR and PB logs (DT features are excluded). An additional forty-two optimized feature selections were generated using the KNN-optimizer routines using the Well A Case 7 dataset as input, which were then ranked considering both accuracy and cost function scores achieved. Table 4 displays the top eight

feature selections based on their performance ranking. Note the optimizer has reduced the fourteen available features to between six and nine for the best performing cases displayed in Supplementary File: Section S3. The three best feature-optimized cases are selected to form Cases 8 (Jaya-D; 8 features), 9 (DE-C; 7 features) and 10 (PSO-C; 8 features).

Those top-three selected cases all include GR0 and PB0 as part of their feature selection. The volatility (GR5 and PB5) and rolling-average volatility (GR6 and PB6) are key selected features in most of the best performing cases (Table 4), together with the rolling average first differential (GR2, PB2). On the other hand, features GR1, GR3, GR4, PB1, PB3 and PB4 are rarely selected by the best performing feature selections derived from the Case 7 dataset.

Case 11 involves all 14 features relating to the GR and DT logs (PB features are excluded). An additional forty-two optimized feature selections were generated using the KNN-optimizer routines using the Well A Case 11 dataset as input, which were then ranked considering both accuracy and cost function scores achieved. The top eight feature selections based on their performance ranking are displayed in Supplementary File: Section S3. Note the optimizer has reduced the fourteen available features to between five and nine for the best performing GR-DT attribute cases. The three best feature-optimized cases, based on accuracy, are selected to form Cases 12 (Jaya-A; 5 features), 13 (CSO-G; 7 features) and 14 (DE-E; 9 features).

Those top-three selected cases all include GR0 and DT0

Table 3. Feature selections (Marked by “×”) for GR-PB-DT recorded logs plus attribute combinations applied to Well A data based on a 0.75 (training): 0.25 (validation) split of data records using the KNN prediction model with various optimizers.

	SCA-G	DE-D	Jaya-C	SCA-F	PSO-A	DE-B	PSO-C	CSO-F
GR0	×	×	×	×		×	×	×
GR1								
GR2	×	×	×	×	×	×		×
GR3				×				
GR4								
GR5	×	×	×	×	×			
GR6	×	×		×		×	×	×
DT0	×	×	×	×	×	×	×	×
DT1		×						
DT2			×			×		
DT3								
DT4								
DT5	×	×	×	×	×		×	×
DT6		×	×	×	×	×	×	×
PB0		×		×	×	×	×	×
PB1		×		×	×			
PB2	×	×			×		×	×
PB3								
PB4			×		×			
PB5		×		×	×	×	×	×
PB6	×	×	×		×			×
Features selected	8	13	10	10	11	8	8	9
Accuracy	0.9980	1.0000	0.9959	0.9959	0.9959	0.9939	0.9939	0.9918
Population size (N)	50	50	40	50	100	50	30	20
Number of iterations	100	100	100	100	100	100	100	100
Fitness score	0.0058	0.00619	0.0088	0.0088	0.0093	0.0099	0.0099	0.0124
Execution time (s)	46.487	101.12	85.567	69.795	176.55	82.315	62.214	86.439

as part of their feature selection. The volatility (GR5 and DT5) and rolling-average volatility (GR6 and DT6) are key selected features in most of the best performing cases (Table 5), together with the rolling average first differential (GR2, DT2). On the other hand, features GR1, GR3, GR4, DT1, DT3 and DT4 are rarely selected by the best performing feature selections derived from the Case 11 dataset.

The cases 3 to 6 involving GR, DT and PB evaluated by the four prediction models considering various combination of the three recorded well logs plus their attributes are:

- 1) Case 3 (twenty-one features): GR0 to GR6, DT0 to DT6, PB0 to PB6 (all available feature)
- 2) Case 4 (eight features): GR0, GR2, GR5, GR6, DT0, DT5, PB2, PB6 (SCA-G solution Table 3)
- 3) Case 5 (thirteen features): GR0, GR2, GR5, GR6, DT0,

DT1 DT5, DT6, PB0, PB1, PB2, PB5, PB6 (DE-D solution Table 3)

- 4) Case 6 (ten features): GR0, GR2, GR5, DT0, DT2, DT5, DT6, PB4, PB5 PB6 (Jaya-C solution Table 3)

Cases 7 to 10 involving GR and PB recorded logs plus their attributes:

- 1) Case 7 (fourteen features): GR0 to GR6, PB0 to PB6 (all available features for GR and PB)
- 2) Case 8 (eight features): GR0, GR5, GR6, PB0, PB2, PB3, PB5, PB6 (Jaya-D solution Table 4)
- 3) Case 9 (seven features): GR0, GR2, GR5, PB0, PB2, PB5, PB6 (DE-C solution Table 4)
- 4) Case 10 (eight features): GR0, GR2, GR5, GR6, PB0, PB2, PB5, PB6 (PSO-C solution Table 4)

Table 4. The cross-validation configurations with the most accurate predictions for each of the 15 (Well A) cases.

Case/Model evaluated		Validation fold	Accuracy	
Case	Model		Mean	St. dev
Raw logs (0-2)	KNN	5	0.9123	0.0095
	SVC	4	0.8665	0.0061
	SVC	5	0.8446	0.0115
	SVC	15	0.9928	0.0073
	SVC	10	0.9836	0.0099
	SVC	15	0.9950	0.0052
	SVC	15	0.9892	0.0095
+Attributes (3-14)	XGB	15	0.9609	0.0170
	KNN	15	0.9745	0.0139
	KNN	5	0.9682	0.0083
	KNN	15	0.9744	0.0133
	SVC	10	0.9779	0.0071
	SVC	4	0.9725	0.0077
	SVC	5	0.9725	0.0077
	SVC	5	0.9815	0.0057

Cases 11 to 14 involving GR and DT recorded logs plus their attributes:

- 1) Case 11 (fourteen features): GR0 to GR6, DT0 to DT6 (all available features for GR and DT)
- 2) Case 12 (five features): GR0, GR6, DT0, DT5, DT6 (Jaya-A solution Table 5)
- 3) Case 13 (seven features): GR0, GR2, GR5, DT0, DT2, DT5, DT6 (CSO-G solution Table 5)
- 4) Case 14 (nine features): GR0, GR2, GR3, GR5, GR6, DT0, DT2, DT5, DT6 (DE-E solution Table 5)

Hence, fifteen case in total, with different feature combinations, are evaluated (Cases 0, 1 and 2 involving just the recorded logs; Cases 3 to 14 involving recorded logs plus selected attributes).

4.3 Multi-fold cross-validation analysis of Well A dataset

Four cross-validation configurations (4-fold; 5-fold; 10-fold and 15-fold) were evaluated for each of the fifteen cases (Case0 to Case 14), applying each of the four prediction models (KNN, LogReg, SVC and XGB). The analysis involved two-hundred and forty cross-validation model evaluations, each expressed in terms of mean and standard deviation for three prediction performance metrics: accuracy, MAE and RMSE. The results for all cases based on the standard deviation of accuracy versus mean accuracy are displayed in Fig. 2, distinguishing the values relating to each prediction model (KNN, LogReg, SVC, and XGB).

As to be expected, as a model's mean accuracy decreases

the standard deviation of its accuracy tends to increase (Fig. 2). In the bottom right corner of that graph a substantial number of the best performing models (KNN, XGB and SVC) overlap with mean accuracy of 0.95 or higher (associated with the models involving log attributes). However, about one quarter of the SVC models executed outperform the other models in terms of the mean accuracy they achieve (accuracy ≥ 0.98) with the Well A datasets. On the other hand, the LogReg model does not perform as well as the other models with none of its evaluations exceeding mean accuracy of > 0.93 . All of the model evaluations achieving mean accuracy of < 0.85 relate to Cases 0, 1 and 2, i.e., those cases that do not involve well-log attributes.

Table 4 lists the single, best performing model in terms of the multi-fold cross-validation analysis for each of the fifteen feature-selection cases considered. It compares the mean and standard deviation accuracy those models achieved and the cross-validation fold involved.

The best performing models were selected as the ones that generated the highest mean minus standard deviation accuracy. This value captures both the expected accuracy and the uncertainty in the accuracy value revealed by the cross-validation analysis. For the cases involving recorded logs only (Cases 0 to 2), the mean accuracy is the lowest, and the 4-fold and 5-fold cross-validation KNN and SVC models generate the best predictions. For the cases involving the three recorded logs plus selected attributes (Cases 3 to 6), the mean accuracy is highest (> 0.98), and the 10-fold and 15-fold cross-validation SVC models generate the best predictions. For the cases involving only the GR and PB recorded logs plus selected attributes (Cases 7 to 10), the mean accuracy is high (> 0.96), and the 15-fold cross-validation KNN and XGB models generate the best predictions for three of those four cases (for Case 9 the 5-fold KNN model provides the best predictions). For the cases involving only the GR and DT recorded logs plus selected attributes (Cases 11 to 104), the mean accuracy is high (> 0.97), and the 4-fold and 5-fold cross-validation SVC model generate the best predictions for three of those four cases (for Case 11 the 10-fold SVC model provides the best predictions).

Fig. 3 cross plots standard deviation accuracy versus mean accuracy achieved by the best performing models for each of the feature-selection cases considered. The distinction between the cases involving well-log attributes and those considering only recorded logs is clear.

Cases 7, 8 and 10 (GR and PB recorded logs plus attributes stand out in Fig. 3 due to their relatively high standard deviation accuracy values. The slightly higher standard deviations for these three cases reflects the slightly higher uncertainty associated with models based on just GR and PB information compared to those based on GR, DT, PB or GR and DT information. This finding is consistent with the outcome for Case 2 (GR and PB recorded logs only), which generates the lowest mean accuracy and highest standard deviation of the other cases considered, excluding Cases 7, 8 and 10.

Table 5. Prediction performances compared for cases involving the three recorded well logs, combined with selected attribute features for Cases 3, 4, 5 and 6 for validation subsets involving 488 data records and models trained with subsets involving 1,462 data records, random split, 0.75 : 0.25.

Case	Model	MAE	RMSE	R ²	Accuracy	Total errors	Error %
0	KNN	0.1475	0.5432	0.7027	0.9262	36	7.377
	SVC	0.1803	0.6005	0.6367	0.9098	44	9.016
	XGB	0.1598	0.5654	0.6779	0.9201	39	7.992
3	KNN	0.0615	0.3506	0.8761	0.9693	15	3.074
	SVC	0.0492	0.3136	0.9009	0.9754	12	2.459
	XGB	0.0451	0.3003	0.9092	0.9775	11	2.254
4	KNN	0.0574	0.3388	0.8844	0.9713	14	2.869
	SVC	0.0615	0.3506	0.8761	0.9693	15	3.074
	XGB	0.0369	0.2716	0.9257	0.9816	9	1.844
5	KNN	0.0615	0.3506	0.8761	0.9693	15	3.074
	SVC	0.0492	0.3136	0.9009	0.9754	12	2.459
	XGB	0.0533	0.3264	0.8926	0.9734	13	2.664
6	KNN	0.0533	0.3264	0.8926	0.9734	13	2.664
	SVC	0.0369	0.2716	0.9257	0.9816	9	1.844
	XGB	0.0287	0.2395	0.9422	0.9857	7	1.434

4.4 Prediction performance of randomly generated Well A validation cases

Based on the multi-cross-validation results only three models (KNN, SVC and XGB) were selected for more detailed analysis, on the basis that they outperformed the LogReg model for all fifteen feature-selection cases evaluated. Randomly selected validation subset evaluations were generated for the Well A datasets; one for each feature-selection case, applying the three models, applying 0.25 training : 0.75 validation splits (equivalent to the 4-fold cross-validation data record division sizes), and 0.9333 training: 0.0667 validation splits (equivalent to the 15-fold cross-validation data record division sizes). This resulted in ninety evaluations; thirty for each of the KNN, SVC and XGB models. The accuracy and the error performance of these randomly selected subsets is displayed in the Supplementary File : Section S4. Table 5 specifically compares the accuracy and error performances of the random validation subsets generated for the Well A cases that involve GR, DT and PB well logs and attributes, applying a 0.25 training : 0.75 validation split.

Case 0 (no attributes involved) generates accuracies ranging from 0.91 (SVC) to 0.93 (KNN) associated with numbers of prediction errors varying from 36 to 44 (7% to 9%). On the other hand, Cases 3 to 6 (involving attributes) generate 15 or less prediction errors (accuracies ≥ 0.97), with the XGB model for Case 6 generating only 7 (1.4%) prediction errors. Fig. 4 identifies that the Case 0 models generate substantially inferior predictions than the cases involving attributes. For the feature-selection cases considering GR, DT and PB logs and

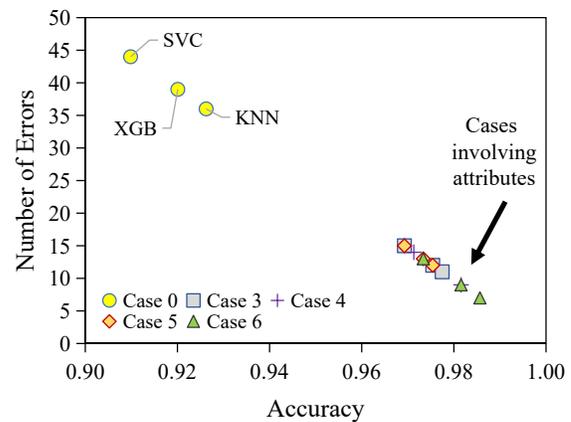


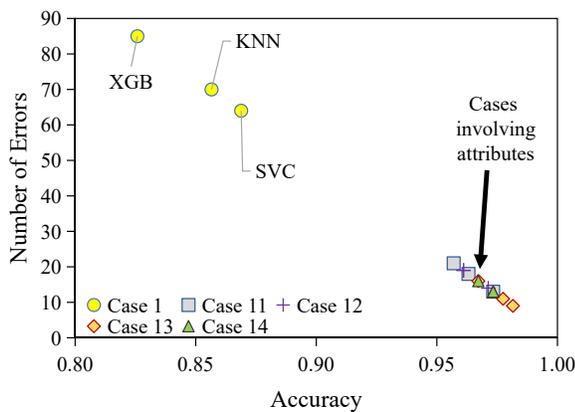
Fig. 4. Prediction results compared for the five cases (for Validation subset with 488 data records) involving the three recorded well logs for the Well A dataset applying a 0.75 : 0.25 training: Validation split evaluated with the KNN, SVC and XGB models.

attributes the Case 6 XGB and SVC models outperform the other cases generating the lowest number of prediction errors.

Table 6 specifically compares the accuracy and error performances of the random validation subsets generated for the Well A cases that involve only the GR and DT well logs and attributes, applying a 0.25 training : 0.75 validation split. Case 1 (no attributes involved) generates accuracies ranging from 0.87 (SVC) to 0.83 (XGB) associated with numbers of prediction errors varying from 64 to 85 (13% to 17%). On the

Table 6. Prediction performances compared for cases involving only the GR and DT well logs, combined with selected attribute features for Cases 10, 11, 12 and 13, random split, 0.75 : 0.25.

Case	Model	MAE	RMSE	R ²	Accuracy	Total errors	Error %
1	KNN	0.2869	0.7575	0.4219	0.8566	70	14.344
	SVC	0.2623	0.7243	0.4715	0.8689	64	13.115
	XGB	0.3484	0.8347	0.2981	0.8258	85	17.418
11	KNN	0.0861	0.4149	0.8266	0.9570	21	4.303
	SVC	0.0738	0.3841	0.8514	0.9631	18	3.689
	XGB	0.0533	0.3264	0.8926	0.9734	13	2.664
12	KNN	0.0574	0.3388	0.8844	0.9713	14	2.875
	SVC	0.0779	0.3946	0.8431	0.9611	19	3.893
	XGB	0.0779	0.3946	0.8431	0.9611	19	3.893
13	KNN	0.0656	0.3621	0.8679	0.9672	16	3.279
	SVC	0.0451	0.3003	0.9092	0.9775	11	2.254
	XGB	0.0369	0.2716	0.9257	0.9816	9	1.844
14	KNN	0.0656	0.3621	0.8679	0.9672	16	3.279
	SVC	0.0533	0.3264	0.8926	0.9734	13	2.669
	XGB	0.0533	0.3264	0.8926	0.9734	13	2.664

**Fig. 5.** Prediction results compared for the five cases (for validation subset with 488 data records) involving only the GR and DT logs for the Well A dataset.

other hand, Cases 11 to 14 (involving GR and DT logs and attributes) generate 21 or less prediction errors (accuracies ≥ 0.96), with the XGB model for Case 13 generating only 9 (2%) prediction errors.

Fig. 5 identifies that the Case 1 models generate substantially inferior predictions than the GR-DT cases involving attributes. For the feature-selection cases considering only GR and DT logs and attributes the Case 13 XGB model outperform the other cases generating only nine prediction errors and achieving an accuracy of 0.9816.

4.5 Applying trained and validated GR-PB Well A models to predict Well B data

As the Well B dataset involves GR and PB well log recordings across the formation boundary it provides the opportunity to independently test models trained and validated with Well A data. Table 7 compares the prediction performances for the KNN, SVC and XGB models applied to Cases 2, 7, 8, 9 and 10 validation subsets (Well A) with testing subsets (Well B). For the validation subsets, Case 2 (no attributes involved) generates accuracies ranging from 0.85 (SVC) to 0.84 (XGB) associated with numbers of prediction errors varying from 72 to 76 ($\sim 15\%$ errors). On the other hand, Cases 7 to 10 (involving GR and PB logs and attributes) generate 30 or less prediction errors (accuracies ≥ 0.94), with the KNN model for Case 9 generating only 14 ($\sim 3\%$) prediction errors. For the testing subsets (Well B), Case 2 generates accuracies ranging from 0.93 (SVC) to 0.91 (XGB) associated with numbers of prediction errors varying from 95 to 12 ($\sim 7\%$ to 9% errors). On the other hand, Cases 7 to 10 generate 104 or less prediction errors (accuracies ≥ 0.92), with the SVC model for Case 8 and the XGB model for Case 10 generating only 71 ($\sim 5.5\%$) prediction errors.

Fig. 6 displays the number of errors versus the accuracy achieved for the cases shown in Table 7 relating to the Well A validation subset (Fig. 6(a)) and the Well B independent testing subset (Fig. 6(b)). For the Well A validation subset the models applied to the cases involving attributes substantially outperform the models applied to Case 2. Moreover, the cases that involve a reduced number of selected features also outperform the prediction performance achieved by Case 7

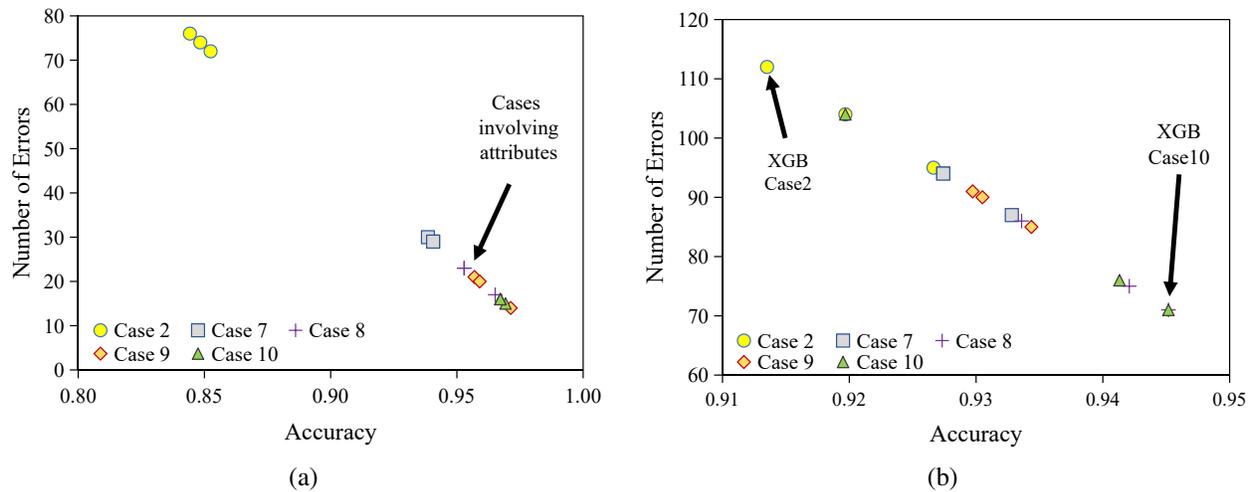


Fig. 6. Prediction results compared for the five cases involving only the GR and PB logs. (a) Models trained and validated with the Well A dataset (for validation subset with 488 data records) and, (b) Well A trained models independently tested with the Well B dataset (for 1,295 data records).

Table 7. Prediction performances compared for cases involving only the GR and PB well logs, combined with selected attribute features for Cases 7, 8, 9 and 10, and independently tested with Well B data, random split, 0.75 : 0.25.

Case	Model	Validation subset Well A (488 Records)						Testing subset Well B (1,295 Records)					
		MAE	RMSE	R ²	Accuracy	Total errors	Error %	MAE	RMSE	R ²	Accuracy (0 to 1)	Total errors	Error %
2	KNN	0.3115	0.7893	0.3724	0.8443	76	15.574	0.1606	0.5668	0.6624	0.9197	104	8.031
	SVC	0.3033	0.7788	0.3889	0.8484	74	15.164	0.1467	0.5417	0.6916	0.9266	95	7.336
	XGB	0.2951	0.7682	0.4054	0.8525	72	14.754	0.1730	0.5882	0.6364	0.9135	112	8.649
7	KNN	0.1189	0.4875	0.7605	0.9406	29	5.943	0.1452	0.5388	0.6949	0.9274	94	7.259
	SVC	0.1230	0.4959	0.7523	0.9385	30	6.148	0.1344	0.5184	0.7176	0.9328	87	6.718
	XGB	0.1189	0.4875	0.7605	0.9406	29	5.943	0.1452	0.5388	0.6949	0.9274	94	7.259
8	KNN	0.0697	0.3733	0.8596	0.9652	17	3.484	0.1328	0.5154	0.7208	0.9336	86	6.641
	SVC	0.0943	0.4342	0.8101	0.9529	23	4.713	0.1097	0.4683	0.7695	0.9452	71	5.483
	XGB	0.0943	0.4342	0.8101	0.9529	23	4.713	0.1158	0.4813	0.7565	0.9421	75	5.792
9	KNN	0.0574	0.3388	0.8844	0.9713	14	2.869	0.1390	0.5272	0.7079	0.9305	90	6.950
	SVC	0.0861	0.4149	0.8266	0.9570	21	4.303	0.1405	0.5302	0.7046	0.9297	91	7.027
	XGB	0.0820	0.4049	0.8348	0.9590	20	4.098	0.1313	0.5124	0.7241	0.9344	85	6.564
10	KNN	0.0615	0.3506	0.8761	0.9693	15	3.074	0.1606	0.5668	0.6624	0.9197	104	8.031
	SVC	0.0656	0.3621	0.8678	0.9672	16	3.279	0.1174	0.4845	0.7533	0.9413	76	5.869
	XGB	0.0656	0.3621	0.8679	0.9672	16	3.279	0.1097	0.4683	0.7695	0.9452	71	5.483

(involving all fourteen available features). This highlights the benefits of optimizing feature selection.

For the Well B testing subset (Fig. 6(b)) the models applied to the cases involving attributes substantially all outperform the prediction performances of the models applied to Case 2, except for the KNN Case 10 model (Table 7). Moreover, most of the cases that involve a reduced number of selected features also outperform the prediction performance achieved by Case 7. Two out of the three models applied to each of Cases 8 and 10 outperform the prediction performances of the Case 9 models. The SVC Case 8 model and the XGB Case 10 models provide the best performance with the Well

B datasets achieving accuracy of 0.945 with only 71 (5.5%) errors. The prediction performance comparison between the Well A validation subsets and the Well B testing subsets are encouraging as they indicate that the models trained and validated with optimized feature selections in data from one well location can be generalized and applied to nearby wells achieving meaningful prediction accuracies.

As well as assessing overall prediction performance of the models it is also important to consider how the prediction errors are distributed between the two formations above and below their point of contact. Confusion matrices provide a useful visual display for providing such information and are

Table 8. Feature importance determined by the XGB model's gain calculation applied to the Well A training results.

Feature	Well log plus attribute feature selection cases														
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
GR0	0.1829	0.2726	0.2923	0.0166	0.0296	0.01845	0.0226	0.0935	0.1245	0.1164	0.1144	0.0214	0.0628	0.0303	0.0242
GR1				0.0068				0.0198				0.0192			
GR2				0.0580	0.0787	0.0655	0.0722	0.0481		0.0644	0.0589	0.0667		0.0681	0.0677
GR3				0.0133				0.0180				0.0154			0.0174
GR4				0.0087				0.0169				0.0142			
GR5				0.0429	0.0687	0.0417	0.0549	0.0510	0.0568	0.0575	0.0488	0.0498		0.0563	0.0578
GR6				0.0313	0.0266	0.0314		0.0357	0.0480		0.0433	0.0323	0.0731		0.0314
DT0	0.4896	0.7274		0.3669	0.5320	0.4797	0.4870					0.4261	0.6069	0.5427	0.4845
DT1				0.0179		0.0103						0.0156			
DT2				0.0310			0.0451					0.0603		0.0712	0.0664
DT3				0.0054								0.0120			
DT4				0.0051								0.0100			
DT5				0.1831	0.1643	0.1712	0.1624					0.1634	0.1666	0.1500	0.1626
DT6				0.0385		0.0330	0.0633					0.0935	0.0906	0.0814	0.0879
PB0	0.3275		0.7077	0.0582		0.0504		0.3705	0.4325	0.4537	0.4289				
PB1				0.0156		0.0079		0.0405							
PB2				0.0199	0.0460	0.0202		0.0520	0.0624	0.0639	0.0568				
PB3				0.0046				0.0179	0.0211						
PB4				0.0027			0.0141	0.0100							
PB5				0.0390		0.0343	0.0404	0.1267	0.1174	0.1397	0.1331				
PB6				0.0346	0.0540	0.0359	0.0380	0.0995	0.1374	0.1043	0.1159				
Total	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

included in Supplementary File: Section S5 (Fig. S4).

The confusion matrices on the left side of Fig. S4 display the error distributions for Case 6 XGB (best performing model involving GR, DT and PB logs and attributes) and Case 13 XGB (best performing model involving only the GR and DT logs and attributes). In both of these cases it is apparent that the models predict the cap rock formation (-1) with perfect accuracy for the Well A validation subset. The few prediction errors generated by those cases are all associated with the reservoir formation (+1). The confusion matrices on the right side of Fig. S4 display the error distributions for Case 10 XGB model (best performing model involving only GR and PB logs and attributes) applied to the Well A validation subset (Figure S4 upper right) and the Well B testing subset (Fig. S4 lower right). For the Case 10 XGB validation subset there are more of the prediction errors located in the cap rock formation than in the reservoir formation. This suggest that the PB log features are not as effective as the DT features in delineating formation (-1). On the other hand, with the Well B testing subset the Case 10 XGB model, trained and validated with Well A data, generates fewer errors associated with formation (-1), with 63 out of the 71 errors generated (~89% of the errors) associated with the reservoir formation.

Illustrating and comparing actual versus predicted formation category for each data record assessed, provides further insight into the location of the prediction errors in relation to the formation contact for the Case 10 XGB model (Supplementary

File: Section S6, Fig. S5). For the Well A validation subset (Supplementary File: Fig. S5A) most of the formation (-1) errors are located close to the formation boundary associated with the transition zone (Supplementary File: Fig. S1) with only four errors positioned elsewhere in formation (-1). The same is true for the Well B testing subset (Supplementary File: Fig. S5B) with only two of the prediction errors positioned some distance from the formation boundary. The few errors that occur at distance from the formation boundary within formation (-1) are associated with isolated higher density bands within that formation (Supplementary File: Figs. S1 and S2) that strongly influence the GR-PB-information based models.

For the Well A validation subset (Supplementary File: Fig. S5A) all of the formation (+1) errors generated by Case 10 XGB are located within 25 m of the formation boundary associated with the thicker shaly layers near the top of the reservoir (Supplementary File: Fig. S1). For the Well B testing subset (Supplementary File: Fig. S5B) all of the prediction errors generated by the Case 10 XGB model are located in four intervals within 60 m of the formation boundary. The intervals containing the errors are all associated with the shaly intervals near the top of the reservoir that are thicker in Well B (Supplementary File: Fig. S2) compared to Well A (Supplementary File: Fig. S1). Hence, the formation (+1) "errors" are useful for identifying shaly layers that characterize the upper portion of this reservoir.

4.6 Feature importance of high performing formation prediction cases

Feature importance of high performing formation prediction cases The XGB model's feature importance "Gini" calculation usefully reveals the relative contribution of each of the selected features to each model's prediction solution. The fractional feature-importance values for the XGB models applied to Cases 0 to 14 are displayed in Table 8.

For Case 0, DT0 is identified as the most important of the three features involved, ahead of PB0 and GR0, in that order. For case 1, DT0 is substantially more important to the XGB model than GR0, as is PB0 for Case 2. In Cases 3 to 6, involving all three recorded logs and their attributes DT0 followed by DT5 (volatility) are the most important features. The GR5 attribute is more important in those solutions than GR0, and the PB6 (moving average volatility) is also assigned more weight than the GR0 feature. In Cases 7 to 10 the PB0 feature followed by the PB5, PB6 and GR0 features are the most important features, making those models heavily dependent on the PB curve features. For Cases 11 to 14 the DT0, DT5 and DT6 features are the most important, and the GR5 and GR6 features are, in most cases, assigned more weight than the GR0 feature. This makes those cases heavily dependent on the DT curve features.

Overall, these feature importance results indicate that in addition to the recorded log values the volatility, and moving-average volatility attributes have substantial influence on the XGB model solutions. This finding is consistent with the correlation coefficients between the well log features and the formation categories (Table 1). The high influence of the volatility (and moving-average volatility) well-log attributes is probable a consequence of the specific characteristics of the two formations considered and would likely not be the case for many other formation boundaries. The cap-rock formation (-1) is highly laminated with recorded log values oscillating across substantial value ranges (Supplementary File: Figs. S1 and S2) over relatively narrow depth intervals, i.e., they display relatively high volatility. This contrasts with the reservoir formation (+1), particularly away from the formation boundary, where the recorded well logs follow more continuous trends over wider depth intervals, i.e., they display relatively low volatility.

5. Discussion

The described technique and case example provided highlight the valuable additional information that well-log attributes bring to bear in the definition of complex formation boundaries in cases where limited well-log information and core analysis is available. This is particularly relevant to many oil- and/or gas-bearing reservoir formations extending across substantial areas within which multiple development wells have been drilled over many years.

In fields for which one, or ideally several wells distributed across the field area exist with a suite of well logs calibrated to core analysis, it is possible to train and validate models based on just two or three recorded well logs combined with selected and optimized attributes. Such trained and validated

models can then be applied to predict formation boundaries more systematically in the large number of other wells for which only two or three suitable well logs are available. The well logs to use and calculate attributes for will vary from field to field based on availability and suitability. GR, DT and PB are a commonly suitable choice as they respond primarily to the lithological characteristics of the formations and they are widely recorded in most wellbores. Neutron and resistivity logs tend to be more problematic for formation characterization purposes, as these logs respond primarily to formation fluid characteristics and tend to vary substantially across the gas-bearing, oil-bearing and water-bearing sections of a reservoir and to water saturation variations through a reservoir. The photoelectric factor log can be usefully exploited as it responds to lithological/mineralogical contrasts in formations, but unfortunately it tends not to be recorded in all wells. The caliper log can also be useful exploited to delineate some formation boundaries because it tends to be influenced, at least in part, by lithological factors. However, the caliper log is also influenced by drilling variables and therefore may vary from well to well due to operational factors.

The technique, as described and implemented, offers the potential to automate formation boundary definition and characterization, especially picking the reservoir tops in a more systematic manner based on just a few available well logs assisted by their attributes. Further studies are required to verify this potential using field-wide studies conducted on well-log data from multiple wellbores.

6. Conclusions

Objective and systematic picking of reservoir formation boundaries with the aid of available well-log data is an important requirement of field development reservoir modelling. However, where the reservoir boundaries are complex and/or transitional, conducting this exercise manually based on limited well-log data across multiple field development wells typically becomes subjective and unreliable. A key constraint is the limited well-log curves recorded in many field development wells. The method proposed in this study, extracting well-log derivative and volatility attributes from two or three recorded well logs, provides additional well-log textural information to better characterize reservoir formation boundaries with the aid of supervised ML algorithms. At a specific formation boundary, this exercise can be configured as a binary classification problem between two formations.

ML prediction performance is improved by applying optimized feature selection to identify and select the most influential recorded well logs and their attributes. Additionally, multi-fold cross-validation analysis of each feature-selected dataset identifies the best data splits to apply to the training / validation datasets to achieve the most reliable formation classification models. Confusion matrices and predicted-versus-actual depth plots assist in identifying the distribution of errors and problematic prediction zones. Trained and validated models supervised with well-log data and the formation boundary defined by core analysis, including selected attributes, from one or several wells can then be deployed to characterize the

formation transition in other field wells.

A case application of the method involving two well sections (Well A with three recorded well logs and Well B with two recorded well logs) transecting a cap-rock (laminated shale) to reservoir (limestone) transitional formation boundary illustrates the implementation of the method. K-nearest neighbor, support vector classification, and extreme gradient boosting models all generate binary formation classifications with high accuracy using just two or three recorded well logs (gamma ray, compressional sonic, and bulk density). In the Well A validation subset (accuracy > 0.97) and the Well B testing subset (accuracy > 0.94) supervised by model solutions trained with Well A data. The best-performing, feature-selected solutions, involving five to ten selected features, substantially improve the formation boundary characterization in the example wells compared to model solutions based only on two or three recorded well logs. Feature importance analysis indicates that for the example formations, the volatility-related attributes have the most influence together with the recorded well logs on the ML model solutions.

Supplementary file

<https://doi.org/10.46690/ager.2023.04.05>

Conflict of interest

The authors declare 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

- Abualigah, L., Diabat, A. *Advances in Sine Cosine Algorithm: A comprehensive survey*. *Artificial Intelligence Review*, 2021, 54(4): 2567-2608.
- Al-AbdulJabbar, A., Elkhatny, S., Mahmoud, M., et al. Predicting formation tops while drilling using artificial intelligence. Paper SPE 192345 Presented at SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Dammam, Saudi Arabia, 23-26 April, 2018.
- Atashnezhad, A., Wood, D. A., Fereidounpour, A., et al. Designing and optimizing deviated wellbore trajectories using novel particle swarm algorithms. *Journal of Natural Gas Science and Engineering*, 2014, 21: 1184-1204.
- Banas, R., McDonald, A., Perkins, T. J. Novel methodology for automation of bad well log data identification and repair. Paper SPWLA 2021-0070 Presented at SPWLA 62nd Annual Logging Symposium, Virtual Event, 17-21 May, 2021.
- Chang, Y., Hsieh, C., Chang, K., et al. Training and testing low-degree polynomial data mappings via linear SVM. *Journal of Machine Learning Research*, 2010, 11(4): 1471-1490.
- Chen, T., Guestrin, C. XGBoost: A scalable tree boosting system. Paper Presented at the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, 13-17 August, 2016.
- Cortes, C., Vapnik, V. Support-Vector Networks. *Machine Learning*, 1995, 20(3): 273-297.
- Cox, D. R. The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1958, 20(2): 215-242.
- Darling, T. *Well Logging and Formation Evaluation*. Houston, USA, Gulf Professional Publishing, 2005.
- Faris, H., Mirjalili, S., Aljarah, I., et al. Salp swarm algorithm: Theory, literature review, and application in extreme learning machines, in *Nature-inspired Optimizers*, edited by S., Song Dong, J., Lewis, A., Cham, pp. 185-199, 2020.
- Fix, E., Hodges Jr., J. L. Discriminatory analysis, nonparametric discrimination: Consistency properties. San Antonio, USAF School of Aviation Medicine, 1951.
- Grant, C. W., Bashore, W. M., Compton, S. Rapid reservoir modeling with automated tops correlation. Paper URTEC 2904037 Presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference, Houston, Texas, 23-25 July, 2018.
- Hong, Y., Kang, C. Automatic well correlation by aligning multiple well using deep neural networks. Paper SPE 202908 Presented at Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, 9-12 November, 2020.
- Lineman, D. J., Mendelson, J. D., Toksoz, M. N. Well-to-well log correlation using knowledge-based systems and dynamic depth warping. Paper SPWLA 1987-UU Presented at the SPWLA 28th Annual Logging Symposium, London, England, 29 June-2 July, 1987.
- Liu, K., Ostadhassan, M. The impact of pore size distribution data presentation format on pore structure interpretation of shales. *Advances in Geo-Energy Research*, 2019, 3(2): 187-197.
- Luo, H., Liu, S., Wang, C., et al. Well-to-seismic calibration in depth domain using dynamic depth warping techniques. Paper Presented at International Geophysical Conference, Beijing, China, 24-27 April, 2018.
- Luthi, S. M. *Geological Well Logs: Their Use in Reservoir Modeling*. Berlin, Germany, Springer Science & Business Media, 2011.
- Rafiei, Y., Motie, M. Improved reservoir characterization by employing hydraulic flow unit classification in one of Iranian carbonate reservoirs. *Advances in Geo-Energy Research*, 2019, 3(3): 277-286.
- Rao, R. V. Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. *International Journal of Industrial Engineering Computations*, 2016, 7: 19-34.
- SciKit Learn. *Supervised and unsupervised machine learning models in Python*. 2023a.
- SciKit Learn. *GridSearchCV: Exhaustive search over specified parameter values for an estimator in Python*. 2023b.
- SciKit Learn. *Bayesian optimization of hyperparameters in Python*. 2023c.
- SciKit Learn. *Cross-validation: evaluating estimator performance*. 2023d.

- Serra, O. *Fundamentals of Well-Log Interpretation-2. The Interpretation of Logging Data*, Developments in Petroleum Science. Amsterdam, Netherlands, Elsevier, 1986.
- Shahid, R., Bertazzon, S., Knudtson, M. L., et al. Comparison of distance measures in spatial analytical modeling for health service planning. *BMC Health Services Research*, 2009, 9: 200.
- Wood, D. A. Hybrid cuckoo search optimization algorithms applied to complex wellbore trajectories aided by dynamic, chaos-enhanced, fat-tailed distribution sampling and metaheuristic profiling. *Journal of Natural Gas Science and Engineering*, 2016, 34: 236-252.
- 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*, 2022a, 6(1): 69-85.
- Wood, D. A. Enhancing lithofacies machine learning predictions with gamma-ray attributes for boreholes with limited diversity of recorded well logs. *Artificial Intelligence in Geosciences*, 2022b, 2: 148-164.
- Wood, D. A. Carbonate/siliciclastic lithofacies classification aided by well-log derivative, volatility and sequence boundary attributes combined with machine learning. *Earth Science Informatics*, 2022c, 15(3): 1699-1721.
- Wood, D. A. Optimized feature selection assists lithofacies machine learning with sparse well log data combined with calculated attributes in a gradational fluvial sequence. *Artificial Intelligence in Geosciences*, 2022d, 3: 132-147.
- Zhang, Y., Li, Y., Guo, W., et al. Differential evolution and the influencing factors of low-maturity terrestrial shale with different types of kerogen: A case study of a Jurassic shale from the northern margin of Qaidam Basin, China. *International Journal of Coal Geology*, 2020, 230: 103591.
- Zhang, B., Ye, T., Xiao, Y., et al. Automated well top picking and reservoir property analysis of the Belly River Formation of the Western Canada Sedimentary Basin. Paper URTEC 3719133 Presented at SPE/AAPG/SEG Unconventional Resources Technology Conference, Houston, Texas, 20-22 June, 2022.
- Zhang, F., Zhang, C. Evaluating the potential of carbonate sub-facies classification using NMR longitudinal over transverse relaxation time ratio. *Advances in Geo-Energy Research*, 2021, 5(1): 87-103.
- Zoraster, S., Paruchuri, Ramoj, P., Darby, S. Curve alignment for well-to-well log correlation. Paper SPE 90471 Presented at SPE Annual Technical Conference and Exhibition, Houston, Texas, 26-29 September, 2004.