

Original article

Predicting brittleness indices of prospective shale formations from sparse well-log suites assisted by derivative and volatility attributes

David A. Wood[✉]*

DWA Energy Limited, Lincoln, United Kingdom

Keywords:

Well-log attributes
brittleness index predictions
feature selection
multi-K-fold cross validation
comparative influence analysis
sparsely logged wells

Cited as:

Wood, D. A. Predicting brittleness indices of prospective shale formations from sparse well-log suites assisted by derivative and volatility attributes. *Advances in Geo-Energy Research*, 2022, 6(4): 334-346.
<https://doi.org/10.46690/ager.2022.04.08>

Abstract:

A technique is proposed that calculates derivative and volatility attributes from just a few well log curves to assist in brittleness index predictions from sparse well-log datasets with machine learning methods. Six well-log attributes are calculated for selected recorded well logs: the first derivative, the moving average of the first derivative, the second derivative, the logarithm of the instantaneous volatility, the standard deviation of volatility, and the moving average of volatility. These attributes make it possible to extrapolate brittleness index calibrations from the few cored and comprehensively logged wells to surrounding wells in which only minimal well-log suites are recorded. Data from two cored wells penetrating the lower Barnett Shale with distinct lithology and five well logs recorded are used to demonstrate the technique. Based on multi-K-fold cross validation analysis, the data matching K-nearest neighbour machine learning model provides the most accurate brittleness index predictions, closely followed by tree-ensemble models. For this dataset, recorded data from three well logs plus calculated attributes matches the brittleness index prediction accuracy that is achieved by the five recorded logs. Moreover, any one of the logs plus their calculated attributes yields better brittleness index prediction performance than that achieved by a combination of just those three recorded well logs. Analysis of the Gini indices of the tree-ensemble models reveals the relative influences of the recorded logs and their attributes on the brittleness index prediction solutions. Such information is used to perform feature selection to optimize the well-log attributes involved to generate reliable brittleness index predictions.

1. Introduction

Limited availability of multiple recorded well logs through gas- or oil-rich shales restrict the ability to accurately identify the most brittle zones within those formations that are likely to respond most effectively to fracture stimulation. For commercial viability purposes, brittleness assessments of shale formations are required to complement kerogen content, type and maturity analysis (Boyer et al., 2006). That combination of analysis makes it possible to zoom in on the formation zones most likely to respond with high oil/gas production rates and resource recovery when stimulated by multi-stage hydraulic stimulation (Grieser and Bray, 2007). As “brittleness” is not a defined geomechanical rock property, a brittleness index (BI)

may be defined in several ways (Yang et al., 2013) with the intention of determining how a formation will likely deform (plastically or elastically) when subjected to increasing stress fields. In recent years, the focus of defining BI (Herwanger et al., 2015) have centred on determining “fracability” of tight formations and quantifying how formations are likely to respond to fracture stimulation (Mews et al., 2019).

Two distinct approaches to establishing BI involve using either geomechanical (Rickman et al., 2008) or mineralogical analysis (Jarvie et al., 2007). The geomechanical methods involve estimating Young’s modulus and Poisson’s ratio (Guo et al., 2012; Gholami et al., 2016; Zhang et al., 2016; Wood and Hazra, 2017) using information either from a shear-wave

acoustic well logs, or by experimentally deriving complete stress-strain curves (Altindag, 2003; Ye et al., 2022), or by extracting attributes from recorded seismic surveys (Mlella et al., 2020; Ore and Gao, 2021). Brittle formations are characterized by displaying low Poisson's ratio and high Young's modulus values. As not many wells record shear-wave sonic logs, and laboratory analysis and seismic attribute determinations are time consuming and costly, obtaining reliable geomechanical measurements for BI determination is a challenge for many wells.

BI based on mineralogy, requiring relatively expensive core analysis or spectral gamma ray and/or nuclear magnetic resonance log data (Wang, 2013), was initially defined using quartz (Qz), limestone (Ls; including calcite veins) and clay (Cy) fractions (Jarvie et al., 2007), as defined by Eq. (1) and expressed as a volume fraction on a scale of 0 to 1.

$$BI = \frac{Qz}{Qz + Ls + Cy} \quad (1)$$

This simplistic BI can be usefully extended to include total organic carbon (TOC) (Wood, 2022a) as a ductile fraction and to distinguish dolomite (DI) as a brittle fraction (Wang and Gale, 2009) as expressed by Eq. (2).

$$BI = \frac{Qz + DI}{Qz + DI + Ls + Cy + TOC} \quad (2)$$

For some formations it is appropriate to distinguish calcite (Glorioso and Rattia, 2012), feldspar and/or pyrite (Jin et al., 2014; Alzahabi et al., 2015) as a brittle components in a BI. As all the methods described to define BI require costly measurements to determine them precisely, it is not feasible to do so in a large number of wells. However, once a BI is calibrated by such direct measurements in one well there is the potential to calibrate that measured BI with well-log data and then use well logs in nearby wells to estimate BI at low cost. Although such calibrations can be based on empirical relationships between BI and certain well logs (Jin et al., 2014), such approaches and multi-linear regression (MLR) are not readily transferable to other basins or other parts of the same basin.

Various machine learning (ML) methods offer more generalizable approaches to well-log based BI predictions (Kaunda and Asbury, 2016; Wood, 2021). ML methods tend to exploit multiple combinations of the basic suite of recorded well logs (Shi et al., 2016; Verma et al., 2016), typically, gamma ray (GR), bulk density (PB), deep resistivity (RS), neutron porosity (NP), and compressional sonic (DT). However, a substantial constraint to applying such ML methods in practice is that many of the wells drilled in shale provinces do not acquire a complete basic suite of well logs. Using four or less well logs as independent variables for ML models generally results in poor BI prediction accuracy. Hence, there is a need to develop ML techniques that can provide accurate BI predictions from sparse suites of well logs. Such methods would open up the possibility to predict BI in many more wells surrounding those few wells with direct BI measurements.

This study proposes and evaluates a method to predict BI from a sparse suite of well logs with the novel approach

of calculating well log derivative and volatility attributes. It uses those attributes together with just 1 to 3 recorded well logs to predict BI. A method involving GR log derivatives and volatility attributes has been recently described and successfully applied for lithofacies analysis (Wood, 2022b). That method is expanded to calculate such attributes from GR, PB and DT logs and apply them to a well-studied dataset of two Barnett Shale wells (Texas, U.S.A.) (Verma et al., 2016; Wood, 2021). to evaluate their usefulness in predicting BI (Eq. (2)), using a minimum number of recorded well logs. Two MLR and five ML BI prediction models are developed and evaluated with a mult-K-fold cross validation to reveal prediction accuracy and uncertainty. Three of the ML models involve tree-ensemble algorithms, and these are used to reveal the influences of key variables on optimum BI predictions involving various combinations of recorded well logs and their calculated attributes as input variables.

2. Materials and methods

2.1 Lower Barnett Shale geological setting

The birth of the modern "shale era" began in the Mississippian-aged Barnett Shale formation in the 1990s. This organic-rich black shale is distributed across the Bend Arch-Fort Worth Basin, Texas (U.S.A.) and produced its first gas in the 1980s. Its burial history, lithofacies and tectonic influences are well understood (Walper, 1982; Abouelresh and Slatt, 2012) and have helped to locate shale-gas production sweet spots within it. Also, the distribution of organic-rich and brittle layers within it have helped to better define its gas distribution (Jarvie et al., 2007). It is generally quartz-rich with limestone/dolomite bands (Singh et al., 2008) which helps to provide it with excellent brittleness characteristics, which respond well to hydraulic fracture stimulation. However, its thickness and quality vary substantially across the basin. It is thickest (>1000 ft/300 m) towards the northeast, where it incorporates more limestone-rich layers. However, it thins westward and south-westward (to < 50 ft/16 m in places) across the Bend Arch and onto the Chappel shelf (Pollastro et al., 2007).

2.2 Lower Barnett Shale well dataset

Published data from two wells (A and B) that penetrate the Lower Barnett Shale (LBS) (Verma et al., 2016) is used. The LBS in Well A extends between 8000 ft and 8470 ft (470 ft thick), whereas the LBS in Well B is somewhat shallower extending between 6490 ft and 6790 ft (300 ft thick). The wells are positioned, some distance apart (~30 miles/48 km) towards the north-eastern margin of the Fort Worth Basin, southwest of the Muenster Arch. Five well logs (GR, PB, RS, NP, DT) were recorded in each well together with extensive cores. Mineralogical measurements, distinguishing Cy, DI, Ls and Qz, have been transformed into continuous mineralogical depth profiles (Verma et al., 2016) facilitating the calculation of mineralogical BI curves. Well A and B logs and BI curves versus depth were subsequently expressed using 1000 sample

Table 1. Pearson and spearman correlation coefficients between the recorded well log and the calculated (Wang and Gale, 2009).

Recorded well logs	Well A		Well B	
	Pearson	Spearman	Pearson	Spearman
GR0	-0.0957	-0.0894	0.0956	-0.0155
PB0	-0.4863	-0.5742	-0.0988	-0.1025
RS0	0.2123	0.5181	0.3734	0.3719
NP0	-0.2516	-0.2836	-0.3438	-0.3636
DT0	0.0493	0.0279	0.1337	0.1487
StH	-0.2565	-0.2443	-0.0211	0.0241

points, and also upscaled by interpolation to 5000 data records for each of the logged intervals to provide about 11 samples/ft for Well A and 17 samples/ft for Well B (Wood, 2021). It is the upscaled well log/BI curve dataset that is evaluated in this study, specifically the GR, PB and DT logs and their calculated attributes. The sampled well log curves (GR0, PB0 and DT0) and mineralogical BI curves calculated with the Jarvie et al. (2007) and Wang and Gale (2009) methods, respectively, are displayed in the Supplementary File (Section S1, Fig. S1 (Well A) and Fig. S2 (Well B)). The Wang and Gale BI curve is the dependent variable evaluated.

Detailed inspection of well-log and BI curve shapes for Wells A and B (Fig. S1 and S2) reveal substantial differences between the two wells, although both wells display substantial BI ranges and high frequency fluctuations. The variability in log values, both within wells and between wells, is a consequence of the distribution of bands of distinct lithology, mainly relatively thin limestone, dolomite, kerogen-rich and clay-rich layers. Whereas, Qz and DI act to enhance brittleness, Cy, Ls and kerogen are more ductile and act to dampen formation brittleness. This can make it worthwhile to assess BI in conjunction with total organic carbon distributions (Wood, 2022a).

2.3 Well-log distributions for Wells A and B

The recorded well log distributions are summarized in statistical terms in the Supplementary File (Section S1, Table S1), together with the stratigraphic height (StH) variable, calculated on a fractional scale from zero (base of LBS in well section) to one (top of LBS in well section). Wood (2021) identified that using the StH as an independent variable improved BI predictions in ML models. The recorded well-log abbreviations used terminate in “0” to distinguish them from the calculated attributes. Although the recorded well-log distributions for PB0 are quite similar in the two wells, the other well logs display substantial differences. GR0 has higher minimum, mean and fifty percentile values in Well B. Although, DT0 displays lower minimum and maximum values in Well A, it has slightly higher mean and fifty percentile values than Well B. This is surprising as Well A is a deeper section than Well B, leading to the expectation that the Well A LBS section should be more compacted with generally lower

DT0 values. This difference is probably a reflection of the distinct lithology distributions in the two LBS sections. RS0 displays a substantially higher mean value in Well B than Well A. Comparing the mean and fifty percentile values of the two wells reveals that the RS0 distribution is asymmetrical and positively skewed in Well A, but almost symmetrical in Well B. Although the NP0 distributions are symmetrical in both wells the mean and fifty percentile values are substantially lower in Well A (~0.18) compared to Well B (~0.21). That NP0 difference between the two wells is more consistent with the expectations for the deeper more compacted LBS section of Well A.

2.4 Pearson and Spearman correlation coefficients for Wells A and B

Comparing Pearson (R) and Spearman (P) correlation coefficients can provide useful insight to distribution relationships. Such comparisons between the recorded well logs and BI for Wells A and B reveals some distinctive relationships (Table 1). R assumes that parametric (~linear) relationships exist between the variables compared. P does not make such assumptions as it considers the ranked positions of the values within the distributions compared, making P more accurate in its assessment of non-parametric relationships (Myers and Sirois, 2004).

There are some substantial differences between P and R values between several of the well logs and BI, particularly in Well A, e.g., PB0 and RS0 (Table 1). The high P and R values for PB0 and BI in Well A distinguish it from Well B. For Well B it is only for GR0 that P and R values diverge significantly, whereas they are in close agreement for Well A. RS0 and NP0 show relatively high P and R values in both wells, and are in close agreement with each other in Well B. The substantial difference P and R values for RS0 in Well A (Table 1) is to be expected based on its highly asymmetrical distribution revealed in Table S1.

Although, the high P and R values for the RS0 and NP0 logs with BI can be usefully exploited by ML prediction models, these relationships are to an extent problematic. RS and NP logs both respond to fluid properties, specifically gas and water saturations and compositions, rather than the bulk properties of the matrix formations. High RS0 and NP0 value shifts are most likely responding to fluid quantities and compositions in pores and natural fractures permeating the formations, rather than the brittleness of the formation itself. Nevertheless, such information is useful from a gas and oil exploration perspective, but it means that lateral and vertical compositional changes in formation fluids can adversely skew BI predictions in models that unduly rely on RS and NP data. For instance, where a swarm of natural fractures crosses a relatively ductile zone, RS0 and NP0 responses may indicate that the zone has a high BI. Conversely, a tight but ductile matrix section with few natural fractures may be identified based on RS and NP responses to have low BI. Although brittle zones are more likely to retain open fractures and microfractures filled with formation fluids than ductile zones, that is not a universal rule that can be relied upon.

2.5 Calculation of well log derivative and volatility attributes

The calculation and use of well-log derivatives and volatility attributes has recently been proposed and applied for ML lithofacies prediction models, initially focused on GR attributes for clastic sequences (Wood, 2022b). This technique has subsequently been more broadly applied, involving GR attributes in assessing more varied lithofacies sequences (Wood, 2022c), and applying other well-log attributes to more complex carbonate/siliciclastic sequences (Wood, 2022d). Here, the technique is further extended to apply attributes for multiple well logs collectively to improve BI predictions in cases where few well-log curves are recorded.

Six attributes are calculated for each well log of interest (GR, PB and DT) using the method already documented for GR (Wood, 2022b). These attributes are defined by the following five equations in which “ L ” refers generically to any well log.

Attribute 1—first derivative ($d'L$):

$$d'L = \frac{L_d - L_{d-1}}{\text{Abs}(d - (d-1))} \quad (3)$$

where L_d is the L value at depth sample point d and L_{d-1} is the L value at depth sample point $d-1$.

Attribute 2—simple moving average (sma) of first derivative ($\text{sma}d'L_{\text{dnw}}$):

$$\text{sma}d'L_{\text{dnw}} = \frac{\sum_{i=1}^{i=\text{nw}} d'L_{d-i}}{\text{nw}} \quad (4)$$

where nw is a specified interval of overlying depth sampling points ($\text{nw} = 10$ for the LBS dataset).

Attribute 3—second derivative ($d''L$):

$$d''L_{\text{dnx}} = \frac{d'L_d - d'L_{d-\text{nx}}}{\text{Abs}(d - (d-\text{nx}))} \quad (5)$$

where nx is a specified interval of overlying depth sampling points over which the second derivative $d''L$ is calculated ($\text{nx} = 10$ for the LBS dataset). There is no reason why the value of nx should be the same as the value of ny , but a value of 10 worked for well for both attributes 2 and 3.

Attribute 4—natural logarithm of instantaneous depth variations ($\ln L_{i(d)}$):

$$\ln L_{i(d)} = \ln \frac{L_d}{L_{d-1}} \quad (6)$$

where $i(d)$ represents the depth sample interval from $d-1$ to d , with d referring to the depth point sample immediately underlying $d-1$.

Attribute 5—standard deviation of attribute 4 for specified time period ($L\sigma_{i(\text{dny})}$):

$$L\sigma_{i(\text{dny})} = \sqrt{\frac{\sum_{i=1}^{i=\text{ny}} (\ln L_{i(d-1)} - \ln L_{id} \text{mean})^2}{\text{ny}-1}} \quad (7)$$

where $L\sigma_{i(\text{dny})}$ is referred to as the volatility for the depth sample interval $d-1$ to $d-\text{ny}$, and ny is a specified depth sample interval ($\text{ny} = 10$ for the LBS dataset).

Attribute 6—simple moving average of L volatility ($\text{sma}L\sigma_{i(\text{dnz})}$):

$$\text{sma}L\sigma_{i(\text{dnz})} = \frac{\sum_{i=1}^{i=\text{nz}} L\sigma_{d-i}}{\text{nz}} \quad (8)$$

where nz is a specified depth sample interval ($\text{nz} = 10$ for the LBS dataset).

The abbreviations GR0, PB0 and DT0 refer to the recorded values of those well logs, numbers 1 to 6 replace 0 in the abbreviations used for the well-log attributes. The well-log attributes used and their abbreviations are summarized as follows:

- GR1, PB1, DT1 for Attribute 1, the first derivative.
- GR2, PB2, DT2 for Attribute 2, the moving average of first derivative.
- GR3, PB3, DT3 for Attribute 3, the second derivative.
- GR4, PB4, DT4 for Attribute 4, the natural logarithm of instantaneous volatility.
- GR5, PB5, DT5 for Attribute 5 the standard deviation of instantaneous volatility.
- GR6, PB6, DT6 for Attribute 6 the moving average of Attribute 5.

2.6 ML and regression models evaluated

Two MLR and five ML methods are applied to model BI predictions using the LBS well log dataset with attributes compiled for Wells A and B. These methods are all well-established and have been previously applied as part of well-log prediction models. Their mathematical formulations are widely discussed in the literature and coded functions to execute them in Python language are publicly available.

MLR models make linear assumptions for the relationships between each independent variable and the dependent variable and exploit classic least-squares regression to minimize prediction errors. In its basic form (referred to as linear regression (LR) here) it employs a coordinate descent optimizer, applies no regularization component in its error computations metric and assumes that there are no dependencies exist between the independent variables. ElasticNet is an MLR method that applies L1 and L2 regularization terms to the residual sums of squares error calculation, acting to minimize the impacts of dependencies and avoid large magnitude coefficients. LR and ElasticNet are the two regression models applied.

Five ML models are employed representing three generic types of algorithms all capable of handling non-linear relationships between independent and dependent variables: data matching, support vector regression, and tree-ensemble regression.

K-nearest neighbour (KNN) is a non-parametric, supervised machine learning algorithm that does not depend on regression relationships and executes rapidly. It establishes the degree of similarity between data records and uses the relative distance between a specified number (K) of the most similar data records to make its predictions.

Support vector regression (SVR) is a non-parametric applying a kernel function to derive the optimum position of a hyperplane in the multi-dimensional hyperspace defined by the variables. When it applies a radial basis function (RBF) as its kernel it is better able to handle non-linear relationships among the variables.

Three tree-ensemble methods are applied:

- Adaptive boosting (ADA) readjusts the weights applied to data records in multiple decision trees in a series of iterations with each training iterations (Freund and Schapire, 1997). ADA assigns higher weights to the data records associated with larger prediction errors in the previous iterations. This forces the model to improve the fit of outlying data.
- Random Forest (RF) evaluates multiple decisions trees applying “bagging” and “out-of-bag” variable selection techniques (Ho, 1998). It makes predictions based on the mean outcomes of all the decision trees evaluated.
- Extreme Gradient Boosting (XGB) applies parallel gradient boosting to multiple trees involving L1 and L2 regularization (Chen and Guestrin, 2016).

The tree-ensemble methods all offer the ability to compare the Gini index values the models assign to the tree branches in their optimum solutions. This makes it possible to establish the relative importance these models assign to each independent variable. This information is used in this study to determine the relative importance of the well logs and log attributes to the tree-ensemble prediction solutions.

The models applied typically require hyperparameters to be set that are dataset dependent. Trial and error, grid search and Bayesian optimisation techniques have been used to determine these hyperparameter values. The key hyperparameter setting values for the seven MLR/ML models applied to the LBS datasets are included in the Supplementary File (Section S2).

Cross validation using sets of randomly selected samples from a multi-variable dataset is a useful tool for deriving statistical measures of uncertainty in prediction modelling. Conducting this on a multi-K-fold basis, with K varying from 4 to 15, also helps to establish the appropriate splits of data records between training and testing subsets to use in order to minimize a model’s prediction uncertainty. A 10-fold cross validation divides (randomly) the dataset into ten equal-sized subsets. Ten separate analysis are then performed using those ten subsets, in each analysis a different one of the subsets becomes the validation subset. The remaining nine subset are then used collectively as the training subset for each of the ten cases run. Means and standard deviations of prediction-error metrics are then computed from the ten analysis run. By comparing the means and standard deviations of 4-fold, 5-fold, 10-fold and 15-fold cross validation the prediction uncertainties associated with a model can be more comprehensively described. Moreover, the splits of data records between training and testing subsets can be selected based on that information to minimize prediction errors and uncertainties. That approach is applied to the LBS dataset for all models evaluated. It is performed by adapting SciKit Learn “RepeatedKFold” and “cross_val_score” algorithms coded in Python.

To ensure that all data records contain the six calculated well-log attributes, the first nineteen data records of each of the 5000-data-record compilations for Well A and Well B are omitted from the analysis. This is a necessary pre-processing step to ensure that Attribute 6 (simple moving average of WL

volatility) can be computed for each data record when $n_y = 10$ and $n_z = 10$ (see Eqs. (7) and (8)). Each well log and attribute is normalized such that its values are distributed on a scale of -1 to $+1$. This is necessary precaution to avoid scaling biases affecting the prediction models. The normalization calculation is described in the Supplementary File (Section S3).

The statistical error-assessment metrics used to monitor and compare BI prediction performance are expressed mean absolute error (MAE), its values expressed on mineral BI scale range of 0 to 1, root mean squared error (RMSE), and coefficient of determination (R^2). The formulas used to calculate these metrics are provided in the Supplementary File (Section S4).

3. Results

3.1 Well-log and log-attribute cases evaluated

Many distinct dataset cases were evaluated for Well A and Well B involving different well-log and log-attribute combinations as independent variables, and with BI as the dependent variable. Of these, the results of ten cases are presented and compared to show the capabilities of well log attributes of the GR, PB and DT logs in predicting BI for these two wells. The independent variables involved in the ten cases are defined as follows.

- Case 0: GR0, PB0, RS0, NP0, DT0, StH (6 independent variables).
- Case 1: DT0, DT1, DT2, DT3, DT4, DT5, DT6 (7 independent variables).
- Case2: GR0, PB0, DT0 (3 independent variables).
- Case3: GR0, PB0, RS0, NP0, DT0 (5 independent variables).
- Case 4: PB0, PB1, PB2, PB3, PB4, PB5, PB6 (7 independent variables).
- Case 5: GR0, GR1, GR2, GR3, GR4, GR5, GR6 (7 independent variables).
- Case 6: GR0, GR1, GR2, GR3, GR4, GR5, GR6, PB0, PB1, PB2, PB3, PB4, PB5, PB6, DT0, DT1, DT2, DT3, DT4, DT5, DT6, StH (22 independent variables).
- Case 7: GR0, GR1, GR2, GR3, GR4, GR5, GR6, PB0, PB1, PB2, PB3, PB4, PB5, PB6, DT0, DT1, DT2, DT3, DT4, DT5, DT6 (21 independent variables).
- Case 8: GR1, GR4, PB0, PB1, PB4, DT0, DT1, DT4, DT5, StH (10 independent variables)
- Case 9: GR1, GR4, PB0, PB1, PB4, DT0, DT1, DT4, DT5 (9 independent variables).

Apart from Case 0 and Case 3, the other cases considered focus upon the recorded well logs GR0, PB0 and DT0 and their attributes, disregarding recorded logs RS0 and NP0. There are two reasons for this: (1) GR0, PB0 and DT0 display lower correlations with BI than RS0 and NP0 enabling their log attributes to better demonstrate their capabilities; and (2) RS0 and NP0 log values are strongly influenced by reservoir fluid compositions and gas/water saturations, which in some shales and tight sediment sequences can distort the relationship between those logs and BI. Specifically, RS0 and NP0 distributions tend to include higher values where natural fractures exist in a sequence, particularly where they are

Table 2. Multi-K-fold analysis results for MLR and ML models applied to the Case 0 Well A dataset.

MAE		4-fold		5-fold		10-fold		15-fold	
		12 cases run		15 cases run		30 cases run		45 cases run	
		Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
MLR	ElasticNet	0.0694	0.00178	0.0694	0.00161	0.0694	0.00272	0.0694	0.00340
	LR	0.0694	0.00178	0.0694	0.00160	0.0693	0.00272	0.0693	0.00340
	ADA	0.0053	0.00023	0.0051	0.00024	0.0048	0.00027	0.0048	0.00032
	KNN	0.0026	0.00032	0.0022	0.00030	0.0015	0.00023	0.0012	0.00024
ML	RF	0.0066	0.00044	0.0061	0.00030	0.0053	0.00034	0.0050	0.00045
	SVR	0.0143	0.00029	0.0141	0.00030	0.0139	0.00034	0.0138	0.00042
	XGB	0.0051	0.00032	0.0050	0.00020	0.0045	0.00028	0.0042	0.00034

Notes: BI predictions from well logs GR0, PB0, RS0, NP0, DT0 plus StH. StDev is standard deviation.

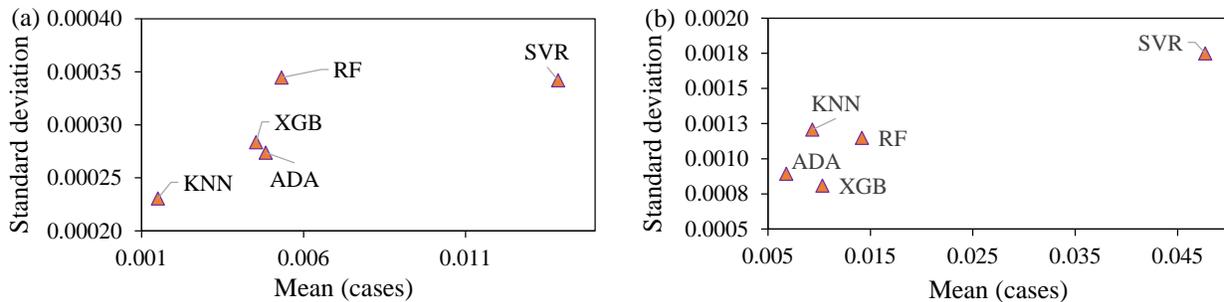


Fig. 1. MAE for 10-fold cross validation from ML model applied to the Well A data for: (a) Case 0, and (b) Case 1.

gas bearing. Although such responses are useful prospectivity indicators they may not reflect the underlying brittleness of the matrix formation. Relationships between GR0, PB0, DT0 plus their attributes and BI are more likely to reflect the underlying brittleness of the formations logged.

3.2 Case 0–five recorded well logs to benchmark BI predictions

Case 0 benchmarks the prediction performances of the five ML and two MLR models with the five available well logs for Barnett Shale Wells A and B. It does so by providing a prediction accuracy benchmark for the models to which the cases involving well log attributes can be compared Table 2 provides the results of multi-K-fold cross validation analysis for each model applied to Well A data with the Case 0 dataset.

The 10-fold and 15-fold cross validation configurations provide the lowest mean MAE values for the ML models. As the MAE standard deviations are slightly lower for the 10-fold compared to the 15-fold configuration, it suggests that a split of 90% training: 10% validation is the optimum choice for the ML models. The MLR models (LR and ElasticNet) deliver much higher mean MAE values than the ML models, but those mean values are almost identical for the four different fold configurations evaluated. For the two MLR models the 5-fold configuration delivers the lowest MAE standard deviations, suggesting that a split of 80% training: 20% validation is the

optimum choice for those two models.

The poorer performance of the MLR models versus the ML models indicates that MLR is not able to adequately exploit the non-linear relationships between the base input logs and BI. For the ML models, the KNN model provides the best BI prediction performance for all K-folds considered (Table 2); mean MAE of 0.0015 representing 0.15% of the BI scale of 0 to 1. The three tree-ensemble methods also provide relatively low and similar prediction errors for Case 0, with the XRB generating a lower mean MAE value than the ADA and RF models (Fig. 1(a)). However, the SVR model does not perform as well as the other ML models for Case 0.

The relative importance of specific logs to the tree-ensemble models can be determined by considering and comparing their Gini Index values. Fig. 1 displays these influences for Case 0 applied to Wells A and B. The ADA and RF models involve quite similar relative influences of the log variables for both wells. For the Well A solutions (Fig. 2(a)), StH and RS0 provide the dominant influences (weights close to 0.3), followed closely by PB0 (weight close to 0.25). However, the other three logs (GR0, NP0, DT0) each provide influence weights of < 0.1.

Overall the influences for the XGB model applied to Well A for Case 0 are similar to those of ADA and RF. However, XGB applies more weight to PB0 and NP0 and less weight to StH than the other two models. The solutions for all three

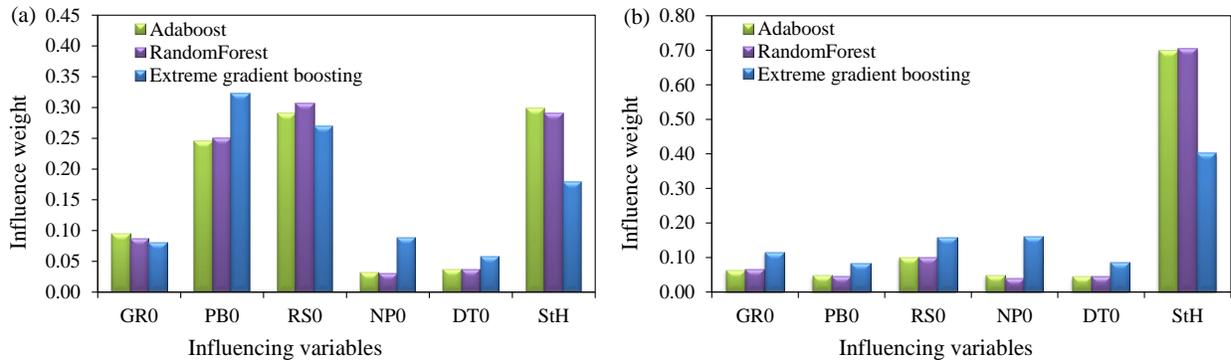


Fig. 2. Relative importance on BI of specific well logs to the tree-ensemble model solutions applied to Case 0 for: (a) Well A, and (b) Well B.

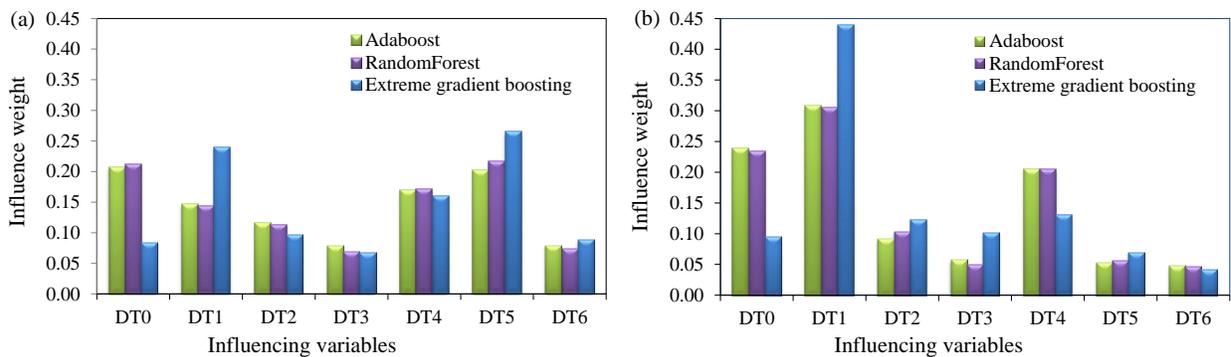


Fig. 3. Relative importance of DT0 and its attributes to the tree-ensemble model solutions applied to Case 1 for: (a) Well A, and (b) Well B.

tree-ensemble models applied to Well B for Case 0 (Fig. 2(b)) are all dominated by StH with weights close to 0.7 for ADA and RF, and ~ 0.4 for XGB. PB0 displays the next highest weight (~ 0.1 for ADA and RF and 0.15 for XGB). However, the XGB solution assigns substantially more weight to NP, GR, DT and PB, in that order, for its Well B solution than the ADA or RF solutions.

3.3 Case 1—a single well log with its attributes to predict BI

Case 1 considers DT0 as the only recorded well log available and combines it with its six attributes (DT1 to DT6) to predict BI. Table S2 (Supplementary File) provides the results of multi-K-fold cross validation analysis applied to Well A data for each model with the Case 1 dataset. As was the case for Case 0, the ML models substantially outperform the MLR models in terms of prediction accuracy and KNN and the tree-ensemble methods also outperform the SVR model. Moreover, the 10-fold and 15-fold cross validation configurations provide the most reliable ML model solutions. Although the mean MAE for the best performing ADA model is substantially higher than the Case 0 model, at 0.0068, it is still only 0.69% of the BI reference scale of 0 to 1. That degree of error can be considered as a relatively reliable model solution based on just one measured well log and its attributes. Moreover, there

is the potential to add variable StH to further improve the Case 1 prediction accuracy.

Fig. 3 displays the relative input variable influences for Case 1 applied to Wells A and B. As was apparent for Case 0, the ADA and RF models involve quite similar relative influences of the log variables for both wells for the Case 1 dataset, and the variable influences for the Well A solutions are quite distinct from the Well B solutions. Again, the influences for the XGB models are distinct from the ADA and RF models. For well A (Fig. 3(a)), the relative influences of the input variables for ADA and RF applied to Case 1 are $DT5 \approx DT0 > DT4 > DT1 > DT2 > DT6 \approx DT3$. In contrast, for XGB (Well A) the sequence of relative influence is $DT5 > DT1 > DT4 > DT2 > DT0 > DT6 > DT3$. Clearly, all six calculated DT attributes are being exploited by the tree-ensemble model solutions, with some given as much or more weight as DT0 in the derived solutions.

For well B (Fig. 3(b)), the relative influences of the input variables for ADA and RF applied to Case 1 are $DT1 > DT0 > DT4 > DT2 > DT5 \approx DT6 \approx DT3$. In contrast, for XGB (Well A) the sequence of relative influence is $DT1 > DT4 > DT2 > DT3 > DT0 > DT5 > DT6$. The relative influences of the DT attributes of the tree-ensemble model Case 1 solutions for both Wells A and B are encouraging because they confirm that all attributes are involved in those

solutions. As will be shown, this is also the position for the other cases considered that involve log attributes, confirming that this technique can be used to generate meaningful BI predictions from just one recorded well log, in this case DT0, plus its attributes.

Many well-log-attribute combinations have been considered for this study, each evaluated with multi-K-fold cross validation, and variable influence analysis conducted. The multi-K-fold analysis results are only shown for all MLR and ML models for Cases 0 and 1 relating to Well A, but they are representative of the other cases considered. The results provided for Case 0 and Case 1 for the MLR and SVR models, clearly reveal the inferior BI prediction accuracy of those models, which is also the outcome when those models are applied to the other cases considered. Hence, the results for those models are not displayed for the other cases considered. Multi-K-fold cross validation analysis is provided only for the best-performing KNN model for the most relevant cases considered (Section 3.4), accompanied by the relative variable influence analysis for the three tree-ensemble models that also generate good BI prediction performance.

3.4 Multi-K-fold cross validation for the KNN model applied to all cases

The multi-K-fold cross validation analysis for all ten cases considered, using separate KNN models for Well A (Table S3, Supplementary File) and Well B (Table S4, Supplementary File), reveal that the 10-fold and 15-fold cases provide the most reliable BI predictions for all cases.

The benchmark Case 0 generates the lowest BI prediction error of the models considered for Wells A and B. However, Cases 6 to 9, involving fewer recorded well logs, also generate BI predictions with very low errors for Wells A and B. It is worthwhile to compare the results for all the cases collectively (Fig. 4) and then evaluate the significance of specific cases. It is apparent from Fig. 4 that Case 2 (just three recorded well logs GR0, PB0 and DT0 with no log attributes involved) generates BI predictions with the most errors compared to the other cases considered. Case 1 (DT only plus attributes), Case 4 (PB only plus attributes) and Case 5 (GR only plus attributes) substantially outperform Case 2 for both Well A and B. This is an important finding as it demonstrates that cases with any one of those well logs plus its attributes can outperform cases with three logs and no attributes.

Cases 0, 3, 6, 7, 8 and 9 all generate BI predictions with small errors making them reliable models for both Well A and B. The deterioration in predictions of Case 3 (five log variables) versus benchmark Case 0 by removing StH as an input is relatively small. This is perhaps surprising as StH has a major influence on Case 0 (Fig. 2) particularly for Well B. It is considered significant that log combination including or excluding StH can provide reliable and competitive BI predictions. Indeed, for some well bores that do not fully penetrate the Lower Barnett Shale Formation or are drilled laterally through the formation without touching its base, StH values cannot be accurately determined. Hence, it is important to construct models that can, if necessary, provide accurate BI

predictions without recourse to the StH variable.

Case 6 involves 22 variables, GR0 to GR6, PB0 to PB6, DT0 to DT6 plus StH, whereas Case 7 involves 21 variables, the same as Case 6 but excluding StH. It is apparent from Tables S4 and S5 and Fig. 6 that Cases 6 and 7 provide almost the same prediction errors for both wells A and B, demonstrating that the GR, PB and DT logs plus attributes can provide good BI predictions with and without the involvement of StH. Moreover, on a mean MAE and standard deviation MAE basis, Cases 6 and 7, for both wells, generate prediction errors comparable to Case 3, which involves the five recorded well logs (GR0, PB0, RS0, NP0, DT0). This is an important finding as it demonstrates that by involving log attributes wells with only three recorded logs can provide BI prediction accuracy comparable to wells with five recorded logs available. The results for Cases 8 and 9 involving customized selections of well logs and attributes will be discussed in Section 3.5, after the justification for feature selection is considered in Section 3.4.

3.5 Relative influences of input variables on tree-ensemble models for Cases 2, 3, 6 and 7

Analysis of the tree-ensemble model solutions reveals for Case 2 (GR0, PB0, DT0) that the ADA and RF solutions display similar influence weights: PB0 (~ 0.50) > DT0 (~ 0.26) > GR0 (~ 0.24) for Well A; and, GR0 (~ 0.37) > DT0 (~ 0.32) > PB0 (~ 0.31) for Well B. On the other hand, the XGB influences for Case 2 are: PB0 (~ 0.44) > DT0 (~ 0.30) > GR0 (~ 0.26) for Well A; and, DT0 (~ 0.37) > GR0 (~ 0.33) > PB0 (~ 0.30) for Well B. For all models, all three variables have substantial influence on the solutions, with PB dominating for Well A and being the least influential for Well B. With only three logs to exploit, clearly, the model needs to incorporate all of them with substantial weight for Case 2, but still are unable to generate highly accurate BI predictions (Fig. 4).

The relative influence analysis for Case 3 (GR0, PB0, RS0, NP0, DT0) reveals that PB0 and RS0 dominate the solutions for Well A, whereas RS0, NP0 and GR0 exert most influence on the Well B solutions. This is displayed in Fig. S3 (Supplementary File). For Well A the relative order of influence is $RS0 \approx PB0 > GR0 > DT0 \approx NP0$. For Well B the relative order of influence is: $NP0 \approx RS0 > GR0 > PB0 \approx DT0$. As already displayed (Fig. 4), this solution is only slightly inferior to benchmark Case 0 that additionally incorporates variable StH.

By removing the influential RS0 and NP0 logs and adding all six attributes for each of the GR, PB and DT logs, Cases 6 (also including StH) and 7 can almost match the BI prediction performance achieved by Case 3 (Fig. 4). Fig. 5 displays the influence analysis for Case 7 (21 variables).

Although Case 7 is a cumbersome one to execute, because it includes so many variables, it provides useful insight to the relative importance of the well-log and attribute variables to the solution, which can be used for feature selection to reduce the number of variables in customized Cases 8 and 9. PB0 (weight > 0.25) dominates all tree-ensemble solutions for Well A Case 7 (Fig. 5(a)), with DT0, DT1, DT2, DT4, DT5, GR0,

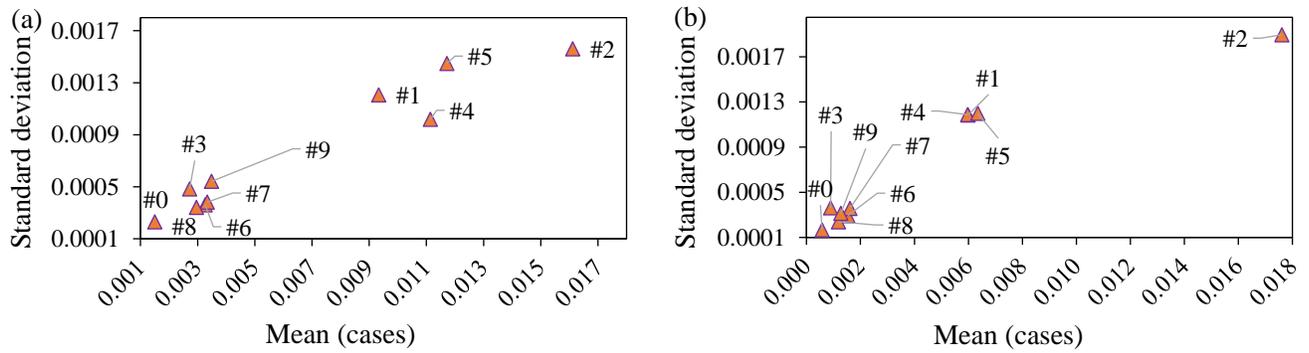


Fig. 4. MAE results for 10-fold cross validation analysis from KNN model applied to Cases 0 to 9 for (a) Well A, and (b) Well B.

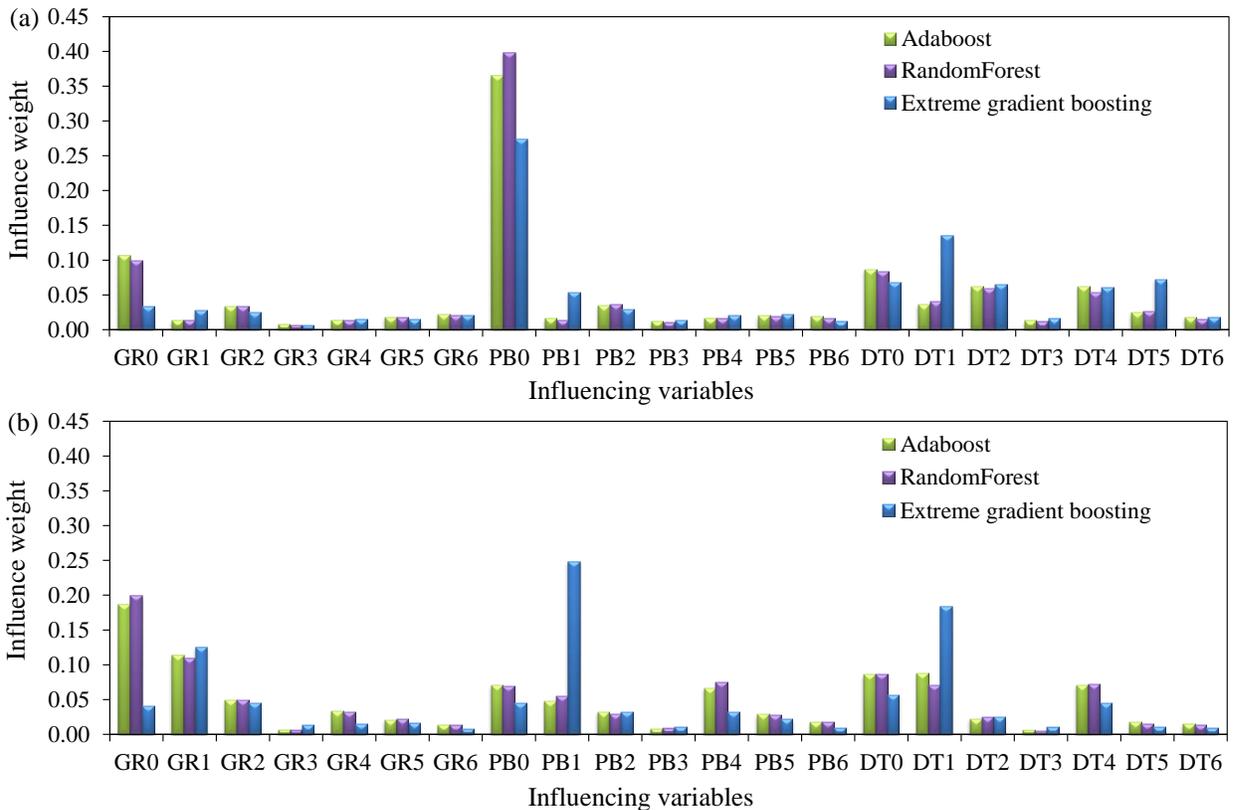


Fig. 5. Relative importance of 21 variables to the tree-ensemble model solutions applied to Case 7 for: (a) Well A, and (b) Well B.

GR2 and PB2 making smaller but greater contributions than the other attributes.

For Well B Case 7 (Fig. 5(b)), GR0, GR1, PB0, PB1, PB4, DT0, DT1 and DT4 exert the most influence. Notably for both wells, the attributes GR3, PB3 and DT3 (second derivative) and GR6, PB6 and DT6 (moving average volatility) exert very little influence on the Case 6 and 7 solutions, justifying their potential removal for feature selection purposes. When StH is added to the Case 7 variable combination to form Case 6, it exerts considerable influence on the solutions for both wells. It rivals PB0 for well A with a weight of ~ 0.30 and dominates

influence for Well B with a weight of ~ 0.7 for the ADA and Rf models and a weight of ~ 0.3 for the XGB model. However, despite the high influence of StH on the Case 6 solutions, the BI prediction performances of Case 6 and 7 are almost identical (Fig. 4).

The results displayed in Fig. 5 indicate that some of the well-log attributes (e.g., the second derivatives: GR3, PB3, DT3) exert very little influence on the tree-ensemble model solutions. It is therefore worth evaluating cases that omit some of the least influential variables (i.e., apply some feature selection) to determine what impact that has on BI prediction

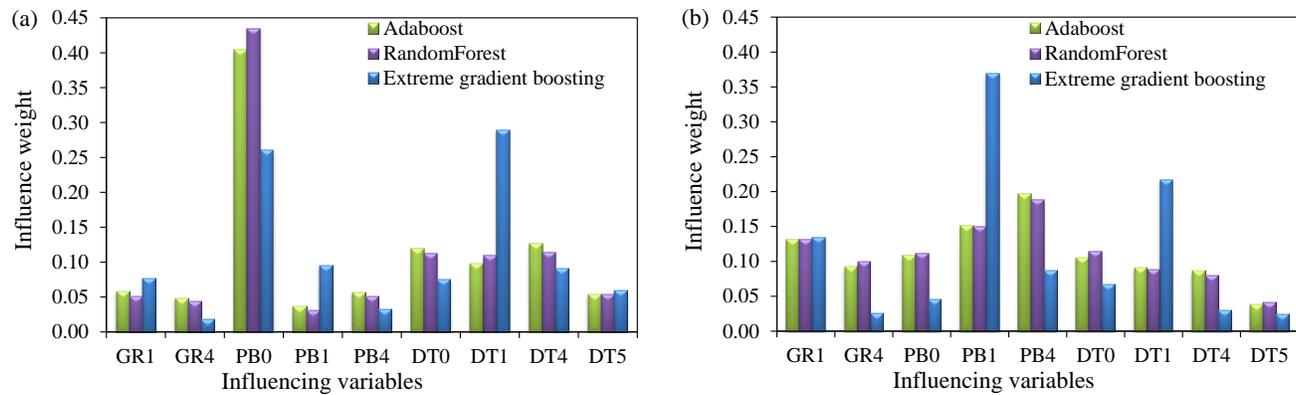


Fig. 6. Relative importance of 9 feature-selected variables to the tree-ensemble model solutions applied to Case 9 for: (a) Well A, and (b) Well B.

performance.

3.6 Feature selected models with 10 and 9 input variables for Cases 8 and 9

The prediction performances and relative variable influences on the model solutions considered so far are used to apply feature selection and compile models using a subset of those included in Cases 6 and 7. Of several feature selections considered Case 8 and 9 show the best BI prediction performance with just 10 and 9 variables, respectively. These cases include the recorded well logs PB0 and DT0 plus the log attributes GR1, GR4, PB1, PB4, DT1, DT4, DT5, forming the 9-variable Case 9, and combined with StH for Case 8. Table S5 (Supplementary File) displays multi-K-fold analysis for KNN and the three tree-ensemble models applied to Cases 8 and 9. All models assessed provide accuracy that rivals that achieved by Case 3, with KNN slightly outperforming the other models for all four K-folds considered.

The influences for Case 8 are illustrated in the Supplementary File (Section S9, Table S6). Variables StH and PB0 exert the dominant influences (weights~0.35) for the Case 8 Well A model solutions, with StH being substantially more influential than other variables for Case 8 Well B. Variables GR1, PB1 and DT1 exert more influence in the XGB model than the ADA and RF models in Case 8 solutions for both wells.

For Case 9, that excludes StH, PB0 exerts the dominant influence (weight~0.4) for the ADA and RF models applied to Well A (Fig. 6(a)). However, for the Well A XGB model, DT1 (weight~0.29) exerts slightly more influence than PB0. On the other hand, PB1 and PB4 exert the most influence on the Case 9 Well B solutions (Fig. 6(b)) for ADA and RF, although influence is generally more evenly spread across the nine variables involved compared to Well A for those two models. For XGB Case 9 Well B, PB1 (weight~0.36) and DT1 (weight~0.21) exert most influence with PB0, GR4, DT4 and DT5 exerting little influence.

Fig. 7 displays a comparison between the prediction performances of Case 2 (GR0, PB0, DT0) and Case 9 (9 feature selected variables including attributes) applying the best performing trained KNN models (based on 90% of the

data records) to randomly selected validation cases (10% of the data records) for Well A and Well B datasets. The prediction accuracy for Case 9 involving feature-selected attributes is substantially improved versus Case 3. Comparisons of the BI prediction performances, in terms of MAE, RMSE and R2, of the KNN models for example validation subsets relating to Cases 0, 2, 3, and 9 are shown in the Supplementary File (Section S7; Table S2). The feature-selected Case 9 solution, based on only the GR, PB and DT recorded well logs plus selected attributes delivers only slightly inferior BI prediction results to those involving 5 recorded well logs.

On a basin scale, spatially and with depth, shale formation petrologic and recorded well-log characteristics tend to vary substantially, as illustrated by Wells A and B, 48 km apart in the LBS. As the characteristics of the shale formation change, so do the relationships of the well log attributes with BI, and their usefulness in contributing to its prediction.

Fig. S5 (Supplementary File) displays Pearson (R) and Spearman (P) correlation coefficients between the well log attributes considered and BI for Wells A and B, to highlight this point. The correlations between the well-log attributes and BI is quite distinct for the two wells considered. These differences have undoubtedly affected the influence of the attributes on the prediction model solutions, as revealed by comparing Figs. 5 and 6 with Fig. S5. For Well A (Fig. S5a), there are substantial differences between P and R values for most of the well logs and well-log attributes versus BI. This suggests that few, if any, of the attribute relationships with BI can be considered as even approximately parametric. This is also the case for Well B (Fig. S5b), but less so, as for PB0, DT0 and most of the DT attributes P and R values are in closer agreement than for Well A.

PB0 displays correlation coefficients with BI of almost -0.6, explaining to some extent why it is a dominant influencing factor in the Case 6 to Case 9 solutions for that well (e.g., Fig. 6(a)). However, DT1 and DT4 display poor correlations with BI for Well A yet they exert substantial influence on the Case 9 solution for that well. The correlation coefficients between the other variables and BI for Well B tend to be low (between +0.05 and -0.05), including that for StH which has dominant

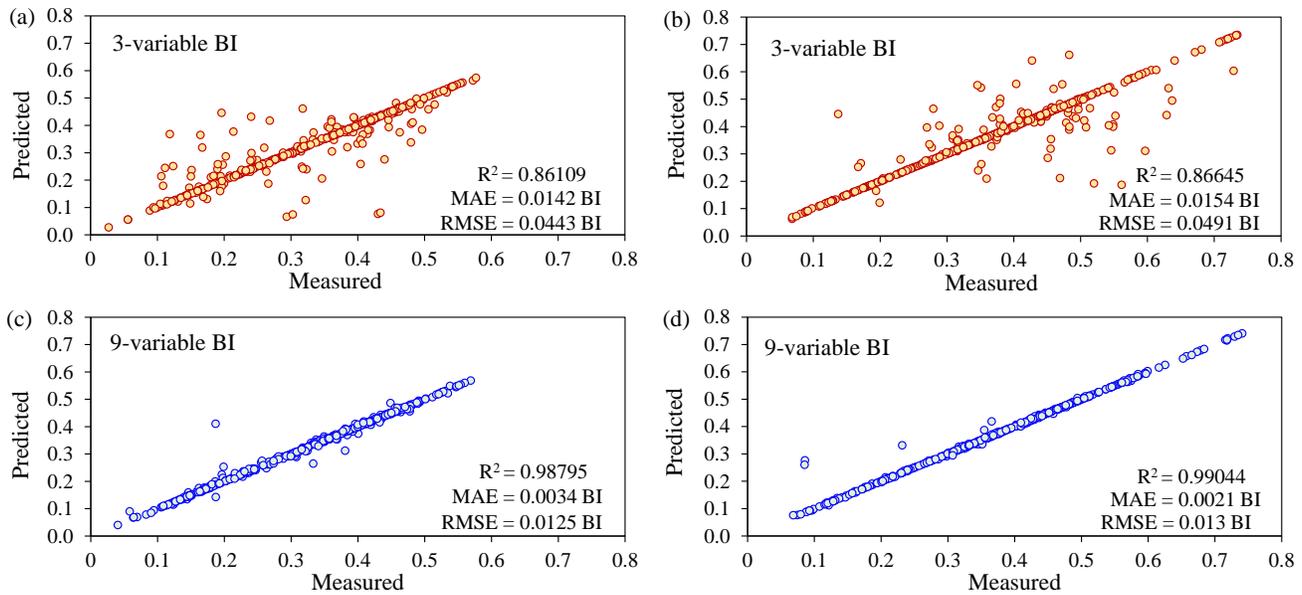


Fig. 7. Predicted versus measured BI for example validation subsets (10% of dataset randomly selected) evaluated by the KNN model for: (a) Case 2 Well A, (b) Case 2 Well B, (c) Case 9 Well A, and (d) Case 9 Well B.

influence on the model solutions for the cases considered that include it for that well (e.g., Case 8, Supplementary File Fig. S4). The combination of moderate negative (PB attributes) and positive (PB attributes) correlations with BI for Well B may explain why the ML models are able to achieve somewhat more accurate BI predictions for Well B versus Well A (as shown in Supplementary File Tables S3 and S4).

4. Discussion

The results presented demonstrate that for the two Lower Barnett Shale wells considered involving well log attributes in the prediction of BI can substantially reduce the number of well logs needed for such predictions. This is an important finding for shale basin exploration and exploitation generally, because the majority of the wells drilled to develop these tight formations tend to record very few well logs. It is not unusual for shale development wells to record only GR, PB and DT logs and, in many cases only over limited depth intervals. The proposed technique opens up the possibility to extract meaningful BI predictions from such wells, provided that there is a nearby well (within ~ 5 to 20 km) penetrating the formation that has been more comprehensively logged (e.g. basic well log suite, plus spectral gamma ray log to provide mineralogical BI measurements and/or shear-wave acoustic log to provide geomechanical BI measurements) and/or cored. In such circumstances, even with just one well log curve available (e.g., Cases 1, 4 and 5 for Wells A and B studied), useful BI assessments can be made. Such BI assessments are important because, in spatially and vertically heterogeneous shale formations, they guide where to optimally position hydraulically induced fractures to achieve the best gas and oil flow rates.

The well log attributes considered (derivatives and volatil-

ity components) are clearly capturing features from the well logs that complement the absolute recorded log values. Essentially, they are capturing the fine detail of curve shape and texture related to the log value variations with depth that can be usefully exploited by ML prediction models. Attributes that are more useful in one formation may be less useful in another. For the two wells studied, the second derivative attribute (GR3, PB3 and DT3) and the moving average volatility attribute (GR6, PB6, DT6) are much less exploited in the prediction model solutions than the other attributes. Further studies are required with other formations to establish whether that is more generally the case.

Although this study is applied to predict a mineralogically-defined BI (i.e., Wang and Gale Eq. (2)), there is no reason why the technique should not also be applied to predict a geomechanically-defined BI from sparse well log data. To do so would require calibration with wells in which shear-wave sonic logs have been recorded, or calibration with extracted seismic attributes from geophysical surveys. Future studies are planned to explore that approach.

5. Conclusions

Six derivative and volatility attributes of well logs are applied to predict BI in shale formations with ML models using a small number of recorded well log curves. Data from two wells (A and B) drilled through the Lower Barnett Shale, in which BI is measured mineralogically from core data, and with a suite of five basic well logs recorded (GR, PB, RS, NP, DT) demonstrate the technique's capabilities. By including attributes for GR, PB and DT, ML models can generate BI predictions from those three well logs with accuracy comparable to that achieved using all five recorded well logs. ML models applied to the calculated attributes of

just one well log (either GR, PB or DT) achieve better BI prediction accuracy than that achieved by using a combination of the three recorded well logs.

KNN model provided the best BI prediction performance for all combinations of recorded logs and their attributes tested. Its prediction performance is almost matched by three tree-ensemble ML methods: adaptive boosting, random forest, and extreme gradient boosting. The tree-ensemble methods provide useful insight regarding the level of influence that each well log and attribute has on the BI prediction solutions, which is effectively exploited for feature selection. Excellent BI predictions are achieved in this way in Wells A and B involving just nine variables (recorded DT plus three DT attributes, recorded PB plus two PB attributes, plus two GR attributes). The first derivative attributes and the instantaneous volatility attributes of GR, PB and DT exert more influence on the BI prediction performances of the tree-ensemble models for Wells A and B than the other attributes considered. This technique is potentially useful in the more widespread estimation of BI in many shale formations for which only a sparse suite of well logs are typically recorded.

Supplementary file

<https://doi.org/10.46690/ager.2022.04.08>

Conflict of interest

The author declares no competing interest.

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

References

- Abouelresh, M., Slatt, R. Lithofacies and sequence stratigraphy of the Barnett Shale in east-central Fort Worth Basin, Texas. *AAPG Bulletin*, 2012, 96 (1): 1-22.
- Altindag, R. Correlation of specific energy with rock brittleness concepts on rock cutting. *Journal of The South African Institute of Mining and Metallurgy*, 2003, 103: 163-172.
- Alzahabi, A., AlQahtani, G., Soliman, M. Y., et al. Fracturability index is a mineralogical index: A new approach for fracturing decision. Paper SPE 178033 Presented at SPE Saudi Arabia Section Annual Technical Symposium and Exhibition, Al-Khobar, Saudi Arabia, 21-23 April, 2015.
- Boyer, C., Kieschnick, J., Rivera, R. S., et al. *Producing Gas from its Source*. New York, USA, Schlumberger, 2006.
- Chen, T. Guestrin, C. XGBoost: A scalable tree boosting system. Paper Presented at Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, August 13-17, 2016.
- Freund, Y., Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 1997, 55: 119-139.
- Gholami, R., Rasouli, V., Sarmadivaleh, M., et al. Brittleness of gas shale reservoirs: A case study from the north Perth basin, Australia. *Journal of Natural Gas Science and Engineering*, 2016, 33: 1244-1259.
- Glorioso, J. C., Rattia, A. Unconventional Reservoirs: Basic Petrophysical Concepts for Shale Gas. Paper SPE 153004 Presented at SPE/EAGE European Unconventional Resources Conference and Exhibition, Vienna, Austria, 20-22 March, 2012.
- Grieser, B., Bray, J. M. Identification of production potential in unconventional reservoirs. Paper SPE 106623 Presented at Production and Operations Symposium, Oklahoma City, Oklahoma, 31 March-3 April, 2007.
- Guo, Z. Q., Li, X. Y., Chapman, M. Correlation of brittleness index with fractures and microstructure in the Barnett Shale. Paper Presented at 74th EAGE Conference & Exhibition, Copenhagen, Denmark, 4-7 June, 2012.
- Herwanger, J. V., Bottrill, A. D., Mildren, S. D. Uses and abuses of the brittleness index with applications to hydraulic stimulation. Paper URTEC 2172545 Presented at Unconventional Resources Technology Conference, San Antonio, Texas, 20-22 July, 2015.
- Ho, T. K. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(8): 832-844.
- Jarvie, D. M., Hill, R. J., Ruble, T. E., et al. Unconventional shale-gas systems: The Mississippian Barnett Shale of North-Central Texas as one model for thermogenic shale-gas assessment. *AAPG Bulletin*, 2007, 91(4): 475-499.
- Jin, X., Shah, S. N., Roegiers, J. C., et al. Fracability evaluation in shale reservoirs-An integrated petrophysics and geomechanics approach. Paper SPE 168589 Presented at Proceeding of SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, 4-6 February, 2014.
- Kuanda, R. B., Asbury, B. Prediction of rock brittleness using nondestructive methods for hard rock tunnelling. *Journal of Rock Mechanics and Geotechnical Engineering*, 2016, 8: 533-540.
- Lai, J., Wang, G., Huang, L., et al. Brittleness index estimation in a tight shaly sandstone reservoir using well logs. *Journal of Natural Gas Science and Engineering*, 2015, 27: 1536-1545.
- Mews, K. S., Alhubail, M. M., Barati, R. G. A review of brittleness index correlations for unconventional tight and ultra-tight reservoirs. *Geosciences*, 2019, 9 (7): 319.
- Mlella, M., Ma, M., Zhang, R., et al. Machine learning for geophysical characterization of brittleness: Tuscaloosa Marine Shale case study. *Interpretation*, 2020, 8(3): T589.
- Myers, L., Sirois M. J. Spearman correlation coefficients, differences between. Oxford, US, Oxford Wiley-Blackwell, 2014.
- Ore, T., Gao, D. Supervised machine learning to predict brittleness using well logs and seismic signal attributes: Methods and application in an unconventional reservoir. Paper SEG 3594773 Presented at SEG/AAPG/SEPM First International Meeting for Applied Geoscience & Energy, Denver, Colorado, 26 September-1 October, 2021.
- Pollastro, R. M., Jarvie D. M., Hill, R. J., et al. Geologic framework of the Mississippian Barnett Shale, Barnett-

- Paleozoic total petroleum system, Bend Arch-Fort Worth Basin, Texas. *AAPG Bulletin*, 2007, 91(4): 405-436.
- Rickman, R., Mullen, M. J., Petre, J. E., et al. A practical use of shale petrophysics for stimulation design optimization: All shale plays are not clones of the Barnett Shale. Paper SPE 115258 Presented at SPE Annual Technical Conference and Exhibition, Denver, Colorado, 21-24 September, 2008.
- Shi, X., Liu, G., Cheng, Y., et al. Brittleness index prediction in shale gas reservoirs based on efficient network models. *Journal of Natural Gas Science and Engineering*, 2016, 35: 673-685.
- Singh, P., Slatt, R. M., Coffey, W. Barnett Shale-Unfolded: Sedimentology, sequence stratigraphy, and regional mapping. *Gulf Coast Association of Geological Societies Transactions*, 2008, 58: 777-795.
- Verma, S., Zhao, T. Marfurt, K. J., et al. Estimation of total organic carbon and brittleness volume. *Interpretation*, 2016, 4 (3): 373-385.
- Walper, J. L. Plate tectonic evolution of the Fort Worth Basin, in C. A. Martin, ed., *Petroleum Geology of the Fort Worth Basin and Bend Arch Area*, 1982: 237-251.
- Wang, F., Gale, J. F. W. Screening criteria for shale-gas systems. *Gulf Coast Association of Geological Societies Transactions*, 2009, 59: 779-793.
- Wang, Y. The method of application of gamma-ray spectral logging data for determining clay mineral content. *Journal of Oil and Gas Technology*, 2013, 35(2): 100-104.
- Wood, D. A., Hazra, B. Characterization of organic-rich shales for petroleum exploration & exploitation: A review-part 3 applied geomechanics, petrophysics and reservoir modeling. *Journal of Earth Science*, 2017, 28(5): 779-803.
- Wood, D. A. Brittleness index predictions from lower Barnett Shale well-log data applying an optimized data matching algorithm at various sampling densities. *Geoscience Frontiers*, 2021, 12(6): 101087.
- Wood, D. A. Assessing the brittleness and total organic carbon of shale formations and their role in identifying optimum zones to fracture stimulate. *Sustainable Geoscience for Natural Gas Subsurface Systems*. Gulf Professional Publishing, 2022a: 129-157.
- 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*, 2022b, 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*, 2022c, 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*, 2022d.
- Ye, Y., Tang, S., Xi, Z., et al. A new method to predict brittleness index for shale gas reservoirs: Insights from well logging data. *Journal of Petroleum Science and Engineering*, 2022, 208: 109431.
- Yang, Y., Sone, H., Hows, A., et al. Comparison of brittleness indices in organic-rich shale formations. Paper ARMA 403 Presented at 47th U.S. Rock Mechanics/Geomechanics Symposium, San Francisco, California, 23-26 June, 2013.
- Zhang, D., Ranjith, P. G., Perera, M. S. A. The brittleness indices used in rock mechanics and their application in shale hydraulic fracturing: A review. *Journal of Petroleum Science and Engineering*, 2016, 143: 158-170.