

Assessment of machine-learning techniques in predicting lithofluid facies logs in hydrocarbon wells

Saba Keynejad¹, Marc L. Sbar¹, and Roy A. Johnson¹

Abstract

Wireline log interpretation is a well-exercised procedure in the oil and gas industry with all its added value from exploration to production stages. It becomes even more important when it is one of only a few available alternatives to compensate for the lack of core samples in a study of lithologic and fluid variations in a well. Yet, as with other purely expert-oriented interpretational techniques, there is always a considerable risk of subjective or technical errors. We have adopted a hybrid approach that links a machine-learning (ML) algorithm to the log interpretation procedure to solve these problems. We have applied this approach to two different hydrocarbon (HC) fields with the aim of predicting the HC-bearing units in the form of lithofluid facies logs at different well locations. The values of these logs are labels of classes that are separated based on their lithologic and fluid content characteristics. After training different MLs on the designed lithofluid facies logs, we chose a bagged-tree algorithm to predict these logs for the target wells due to its superior performance. This algorithm predicted HC units in an accurate interval (above the HC-fluid contact depth), and it showed a very low false discovery rate. The high-accuracy rate, speed of analysis, and its generalization ability, even in data-deficient cases, accentuate why including ML algorithms can improve the understanding of the subsurface at every phase of the exploration and production process. The proposed approach of using ML algorithms, trained and tuned based on the expert's knowledge of the reservoir, can be modified and applied to future wells in a HC field to significantly minimize the risk of false HC discoveries.

Introduction

Boreholes and calibration wells are direct sources of information in the study of lithology and fluid content in a hydrocarbon (HC) reservoir. The cores and drilling reports, where they are available and well documented, provide us with this first-hand observation of the subsurface characteristics. In those cases in which this source is not reliable (e.g., a low core-recovery ratio or a lack of samples at a certain depth) or not available at all, wireline logging can fill the gap to help with interpretation of the lithofacies and potential fluid column at a well location. This interpretation process is not always easy, straightforward, or accurate. In fact, it is usually time consuming, subject to human error, and may be affected by other complicating factors such as poor log calibrations, conflicting log results, and lack of data. A better alternative is to use a hybrid approach that implements a computational method to retrieve as much objective information as possible from the logs while inserting the first-hand information by the analyst to model the vertical lithologic and fluid variations in each well.

One of the quantitative techniques for the first part of this hybrid approach is machine learning (ML). ML al-

gorithms are data-driven techniques that can learn the intended properties within a data set such as classes and trends and then extract those features from unseen data as well. Different techniques usually fall into one of the categories of supervised or unsupervised algorithms.

Supervised ML algorithms build predictive models with known outputs for different observations. These outputs can be continuous values or class labels, for which regression models (function approximators) and classifiers are used, respectively. However, unsupervised algorithms discover some measure of proximity or shared features within the learning set to subdivide the data into clusters, without knowing the desired outputs. Either way, there is always some measure of training performance, the improvement of which facilitates algorithm performance on the unseen data.

Despite the fast-growing number of successful cases published in recent years, application of ML in HC exploration is not yet established as a standard procedure within the exploration and production world. Among different fields of application, ML algorithms have en-

¹University of Arizona, Department of Geosciences, Tucson, Arizona, USA. E-mail: sabaa@email.arizona.edu; msbar@email.arizona.edu; johnson6@email.arizona.edu.

Manuscript received by the Editor 4 July 2018; revised manuscript received 1 October 2018; published ahead of production 01 April 2019; published online 28 May 2019. This paper appears in *Interpretation*, Vol. 7, No. 3 (August 2019); p. SF1–SF13, 11 FIGS., 1 TABLE.

<http://dx.doi.org/10.1190/INT-2018-0115.1>. © 2019 Society of Exploration Geophysicists and American Association of Petroleum Geologists. All rights reserved.

joyed increasing attention in facies-recognition studies in recent years. Artificial neural networks (ANNs) have been used in many cases, for instance, in complex carbonate reservoirs, in which lithofacies recognition based on wireline logs can be very challenging (Qi and Carr, 2006; Al Moqbel and Wang, 2011). Torres and Reverón (2014) and Zhao et al. (2014) use different support vector machine (SVM) approaches for lithofacies classification in reservoir modeling. Other ML algorithms such as generalized boosted regression modeling and quadratic discriminant analysis (DA) have also been successfully tested and applied by Aleardi and Ciabbari (2017) and Al-Mudhafar (2017).

In this study, we are proposing a hybrid approach that merges data-driven (ML) and knowledge-driven methods. The former yields superior computational power, accuracy, and resolution, whereas the interpreter can get around the potential data deficiency and/or acquisition problems through the latter. We have applied this technique to two different data sets to create lithofluid facies logs with the main goal of predicting HC-bearing units. This approach can be modified for different fields depending on their data availabilities and reservoir characteristics.

The first step of this framework includes data preparation, qualitatively and statistically investigating the logs, and then designing target lithofluid facies logs accordingly. In this step, the expert defines the number and type of target classes and chooses the logs. Note that most ML algorithms cannot predict new classes beyond the limited set of classes on which they are trained. For reservoirs in which encountering new facies in unseen wells is expected, a series of neural networks (ART2) with the capability of expanding predicted clusters can be used (Chang et al., 2000). The chosen logs (referred to as features, attributes, or predictors) should demonstrate a desired degree of relevancy to the target classes. The logs and classes will be used to train the algorithm on the wells with known lithofluid facies logs (training wells). After obtaining satisfying results in training, the algorithm can be applied to the unseen wells (target wells) to predict the lithofluid facies classes.

Different ML algorithms were tested on two fields: a self-organizing feature map (SOM) as an unsupervised algorithm, along with a multilayer feed-forward neural network (MLFN) and a bagged-tree (BT) classifier as supervised algorithms. The choice of the supervised or unsupervised ML algorithm in similar cases strongly depends on the data availability and the objectives of the study. The more known data are available (e.g., boreholes), the more reason to incorporate them in training a supervised algorithm. In the case of insufficient data, an unsupervised algorithm can be a better choice to address the uncertainty in the number of clusters. The best approach is to test different ML algorithms and evaluate the results with the known data and the general information about the field before applying one.

We have implemented the same workflow on both fields to better assess the methodologies. All algorithms were trained on each data set, and after evaluating the training results (qualitatively and/or quantitatively), the best algorithm (BT) was applied to the target wells. Depending on the available information from each data set, we then assessed our predictions using a previously interpreted reference in each case.

The available data sets for this analysis include four boreholes from Heidrun Field, off-shore Norway, and eight boreholes from the Kupe Field in the Taranaki Basin, off-shore New Zealand. Heidrun is a producing oil field with associated gas, and Kupe is a gas-condensate field. Both fields are clastic reservoirs, but with distinct properties such as the amount of alternating shale layers and the quality of the reservoir sand. Even though the framework is the same, the parameters of the algorithms for each field were designed separately based on available wireline logs and the facies types. Each data set includes different wireline logs accompanied by drilling information available for some wells. These are the primary sources of information needed by the expert to design and train the algorithm and then interpret the results of its application. According to the objectives of this study, we chose to keep the class labels to a minimum, regardless of the stratigraphic and depositional factors, to highlight the HC-bearing units. To evaluate the methodology and results, we have excluded parts of the available data, either in the form of one whole well and/or a certain depth interval. The excluded parts form our target wells and sections, whereas the rest are used as training wells.

This study shows promising results in applying ML algorithms in HC exploration. SOM can be used to define target clusters in pilot stages of exploration or underexplored areas in which no cored boreholes are available for training. Supervised algorithms, however, could successfully predict HC units in both fields, consistent with fluid contact depths and without false discoveries. Such advantages can help in lowering the risk of overestimating a reservoir's capacity in field development stages.

Heidrun Field

Located on the mid-Norwegian continental shelf, Heidrun Field was formed in an extensional tectonic phase during the Late Jurassic to Early Cretaceous. The Middle Jurassic Fangst Group clastic reservoir was deposited in a shallow marine to fluvial environment. It comprises three mostly clean sandstone formations, with the Not Formation being the more shaly layer that thickens toward the southwest (approximately 19 m in well 6507/7-4). For more on the stratigraphy of the Fangst reservoir and Heidrun Field geology, refer to Harris (1989) and Morton et al. (2009). We focused on a section within four vertical wells that includes the Melke Formation shale (Viking Group, Middle to Upper Jurassic), the Fangst Group,

and the underlying Ror muddy sandstone and Tilje sandstone and shale (Båt Group, Early Jurassic).

To create the training database, we assumed that only well 6507/8-1 has complete drilling information, whereas well 6507/7-4 has partial records. With this assumption, we simulate a case in which the core samples of the second well were lost for a certain depth interval. The excluded section of the second well (2480–2600 m) is, in fact, part of the oil-bearing sandstone in the reservoir. The other two wells (6507/7-3 and 6507/7-8) are assumed to lack any core samples. In this way, we can “mask” the excluded parts of our available information only to use them later for testing the approach and validating the results.

The compiled database includes depths and the associated attribute values extracted from in situ and computed logs. To create the lithofluid facies logs, crossplots of different logs were analyzed. According to drilling information provided by Norwegian Petroleum Directorate (n.d.), four classes are defined as the dominant facies shaping the reservoir: shale (including mudstone and siltstone), brine sand, oil sand, and gas sand. Because well 6507/8-1 is the only well with gas-sand samples, we chose to use it in the training phase, and we evaluate any potential false discovery of gas sands in other wells.

The best separation between classes was seen in the lambda-mu-rho (LMR) analysis on a crossplot of $\lambda\rho$ (LR) versus $\mu\rho$ (MR) logs, noting that incompressibility (λ) and rigidity (μ) are, by definition, pore-fluid and rock-matrix indicators, respectively. [Goodway et al. \(1997\)](#) compare the Lamé constants embedded in the P- and S-wave velocity or in a ratio (V_P/V_S or λ/μ) with LR and MR coefficients, and they demonstrate a significant increase in sensitivity toward petrophysical variations with LR and MR coefficients. The LMR analysis shows promising results in petrophysical discrimination in HC reservoirs. [Young and Tatham \(2007\)](#) identify gas sands by applying LMR inversion on young, unconsolidated sediments. LMR analysis can also become a crucial tool in reservoirs in which a decrease in V_P/V_S occurs without a pore-fluid-related increase in V_S ([Close et al., 2016](#)). The LMR logs were computed using in situ logs (V_P , V_S , and density) as follows:

$$\mu\rho = Z_S^2, \quad \lambda\rho = Z_P^2 - 2Z_S^2,$$

where Z_P and Z_S are the P-wave and S-wave acoustic impedances, respectively.

The results of LMR analysis were used to create the initial lithofluid facies training classes. Each interval was then approved or modified by the reference information to form the class labels for our database. This is the confirmed part of the data set on which the algorithms will be trained, and based on that, the missing information will be recovered/predicted.

The other attributes (logs) were selected by plotting the probabilistic density function (PDF) histogram of each log for each target class. Among all of the wireline

logs that were available and well-recorded within the reservoir interval at all well locations, we interpreted six of them to be good target indicators. These six logs, including LR, MR, porosity (phi), Poisson's ratio (PR), shale volume (V_{Sh}), and P-impedance (Z_P), along with true vertical depth (TVD), form the seven predicting attributes for Heidrun Field. We found that including depth as an attribute can improve the HC-prediction rate because it can act as a constraint in our 1D data sets.

Kupe Field

The Paleocene Farewell Formation is the primary reservoir for the Kupe gas-condensate Field, in the southeastern Taranaki Basin, off-shore New Zealand. The Farewell Formation is composed of medium- to coarse-grained sandstones with interbedded shale layers, deposited in fluvial to coastal braided plains ([Pang and Collen, 1996](#)). Our primary focus in this study is on sections in eight wells that contain the Farewell sandstone reservoir. The reservoir quality of the Farewell Formation can be highly affected in some parts depending on the type of clay minerals within the sandstone ([Martin et al., 1994](#)).

We used four vertical wells (Kupe South-6, Kupe South-7ST1, Kupe South-8, and Momoho-1) to train the ML algorithms for predicting the lithofluid facies in other wells. The training wells, unlike the four target wells, have V_S logs that were used to create LMR and PR logs and crossplots. We also used crossplots of the gamma ray (GR) versus the difference between deep and shallow resistivity logs (RESD and RESS) to design the lithofluid facies target logs. Based on these analyses and the available drilling reports ([New Zealand Petroleum and Minerals, 2017](#)), four classes were defined: shale, brine sand, HC sand, and HC shaly sand. The latter two classes are HC-bearing sandstones with different degrees of shale content as an indicator of reservoir quality. The defined classes can be different in any field depending on the reservoir, data availability, or the objective of the study. For instance, minor coal seams occurring throughout Farewell sandstones in some wells did not improve the overall prediction when included as a separate class, nor did they relate to the objective of this paper.

The lithofluid facies logs were then compared with the existing composite logs so that the defined classes and possible fluid contacts are in agreement with the drilling information. This corrects for any misclassifications due to inaccurate wireline log measurements and/or potentially oversimplified labeling. Based on our investigations of available wireline logs, GR, resistivity, and Z_P logs were selected along with depth values, for predicting lithofluid facies by ML algorithms. Note that to train and apply the algorithms, the choice of predictor logs is limited to the ones that are available in target wells. So the LR and PR logs, which were used in creating the lithofluid facies classes, could not be included in the training because the target wells did not

have V_S (and, consequently, LR and PR) logs. The parallel coordinates plot in Figure 1 shows how each of the selected predicting features can separate the target classes. It also shows the importance of ML algorithms by noticing the challenge of separating overlapping classes in a multidimensional attribute space.

Application of ML algorithms

Machine learning can be thought of as a self-explanatory term, noting that the word “machine” implies a wider spectrum of definitions such as computational modeling. In other words, any data analysis method with the aim of discovering trends and features within a data set and from the data set itself without relying on a predetermined equation is considered a ML algorithm. The algorithm learns the desired information from the data set, which is, in essence, similar to the procedure of learning from experience in human beings.

ANNs are a particular branch of ML, originally inspired by biological neural networks. ANNs are well-known as powerful ML techniques in solving complex and nonlinear problems governing a large amount of multiattribute data sets. Supervised ANNs take in the observations (e.g., logs) and relate them to the associated targets (e.g., known lithofacies) via nonlinear activation functions of their innermost hidden layer(s). However, unsupervised algorithms have the advantage of not depending on the availability of known targets for making predictions. But for the same reason, their re-

sults need to be evaluated more rigorously and validated by reliable evidence before being authorized in a decision-making procedure.

The number of layers, their neurons, the connection weights between layers, and the direction of information flow in an ANN define its architecture, which in turn depends on the problem. The training algorithms are optimization problems that minimize the difference between prediction and the target by tuning the connection weight values accordingly. For a history of the application of ANNs in geophysical studies until 2002, see [Poulton \(2002\)](#).

The bagging or bootstrap aggregating method is generally used to increase stability and reduce the variance of an ML algorithm. This procedure is typically applied to the classification and regression trees (CART), which tend to overfit the training data. CARTs are decision-making models that relate features via “branches” to a “leaf,” which is a predicted value (regression) or label (classification). The branches split at several points based on variations in the features.

The BT procedure begins by creating replicates of the learning set through bootstrap resampling, i.e., random sampling with replacement from the training data set ([Breiman, 1996](#)). This way, each decision tree is trained on a bootstrap sample set, the outcome of which is an aggregation of the results through voting or averaging for classification and regression problems, respectively.

Artificial neural networks

SOM is an unsupervised ANN that clusters data while preserving their topological relationships. SOM uses the Kohonen rule in a competitive layer, by which the winning neurons are determined based on their “closeness” to the prototype or initial vectors ([Kohonen, 1987](#)). In each iteration, the connection weights to the winning neurons will be updated, putting the neurons with similar features in one cluster. SOM entered the seismic interpretation field mainly as a tool in horizon tracking and waveform recognition, and it is still being used in facies mapping studies in more innovative workflows ([Liu et al., 2017](#); [Zhao et al., 2017](#)).

We used SOM as an initial and alternative data-driven method to estimate the number of clusters, independently from the LMR-based clustering results. For this purpose, we measured the silhouette parameter (as defined below) to evaluate the optimum number of SOM-driven clusters. This parameter is calculated as a ratio for each point, using a measure of dissimilarity (e.g., Euclidean distance, correlation, cosine of two vectors, etc.), as follows:

$$S_i = \frac{B_i - A_i}{\max(A_i, B_i)},$$

where A_i is the average distance between point i and the rest of the points in the same cluster and B_i is

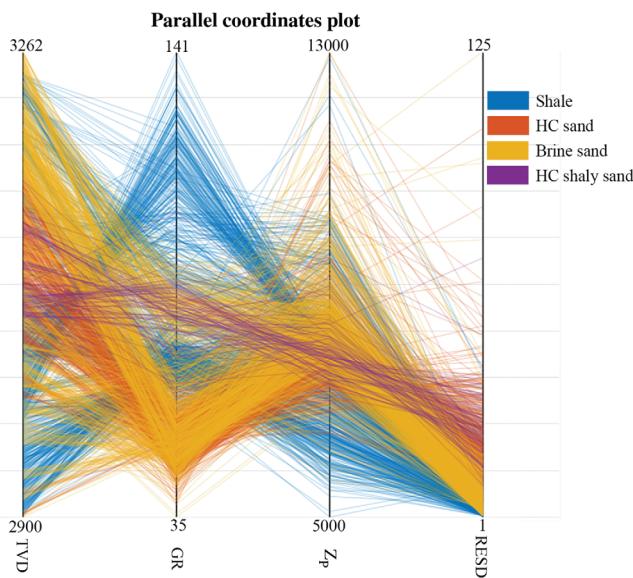


Figure 1. The lithofluid classes plotted on predicting-feature coordinates (parallel coordinates) for New Zealand data: true vertical depth (TVD), gamma ray (GR), P-impedance (Z_p), and deep resistivity (RESD). Each feature is scaled over its range to better show the separation of classes. Also note how the GR (as an example) separates some of the shales with a higher GR from the other lithologies, but other shales with lower GR are similar in value with the brine sands. Most of these low-GR shales, however, can be separated from the brine sands on the TVD or Z_p coordinates.

the distance between that point and the points in the closest cluster (Rousseeuw, 1987).

After running SOM for feature maps with three, four, five, and six neurons (each neuron representing a cluster) and calculating the averages of the S_i in each case, the optimum number of clusters determined by the silhouette parameter was four for both fields (Figure 2). We then applied the SOM network trained for a four-class scenario on each data set. Because of the unsupervised nature of this network, we must label its predicted clusters after it has been applied based on the available log and other data.

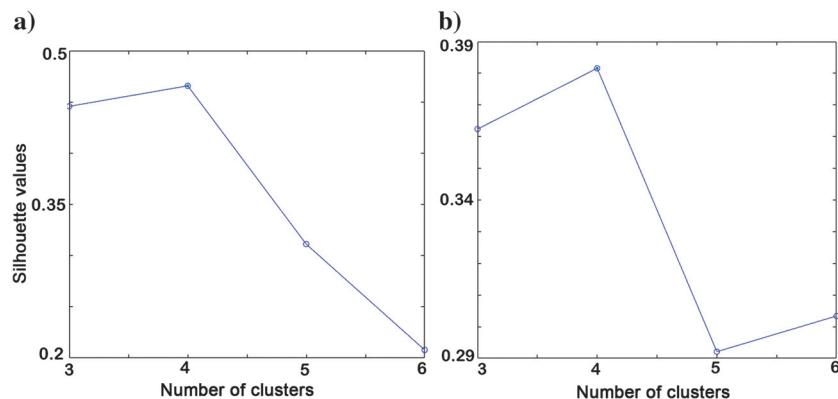
MLFN is one of the common supervised ANNs that can be used in classification and function-approximation problems. It has at least one hidden layer and updates the connection weights and biases in a backward direction (backpropagation) using one of the many gradient-based training algorithms available. The output layer uses either a sigmoid or a linear transfer function for classification and regression problems, respectively.

We used a single hidden layered MLFN with 10 neurons as a classifier to predict lithofluid facies logs in the two fields separately. The network uses 15% of the training data, randomly selected, for validating the results. The validation value is a generalization measure that estimates the algorithm's capability in reproducing what is learned in the new "unseen" data set.

Bagged trees

We used a random-forest approach to grow the BTs for the BT algorithm. This means that, in addition to the bootstrap sampling of the training data set for each tree, the features (log values) at each split are also randomly sampled without replacement (Breiman, 2001). The training procedure can be summarized as follows:

- 1) Take a bootstrapped sample of the training set to form a tree.
- 2) Form a splitting node by randomly sampling features (e.g., splitting a branch into $PR > 0.3$ and $PR < 0.3$).
- 3) Repeat step 2 at each split in the tree until grown as large as desired.
- 4) Repeat steps 1–3 for a large number of trees.



- 5) Over all the trees, count the number of predicted classes for a specific observation (in BT, a set of all the splitting nodes on a branch that leads to a certain leaf is called an observation).
- 6) The majority vote of an observation determines the class assigned to it.

Brieman (2001) shows that the combination of voting over a large number of trees and random sampling of features results in reduction of variance and bias. Thus, there is no need for pruning the trees because the overfitting problem is already addressed by these two qualities. Figure 3 shows a small part of one of the trained trees as an example.

We tested different ML algorithms (including SVM and DA) with a fivefold cross-validation factor in an ini-

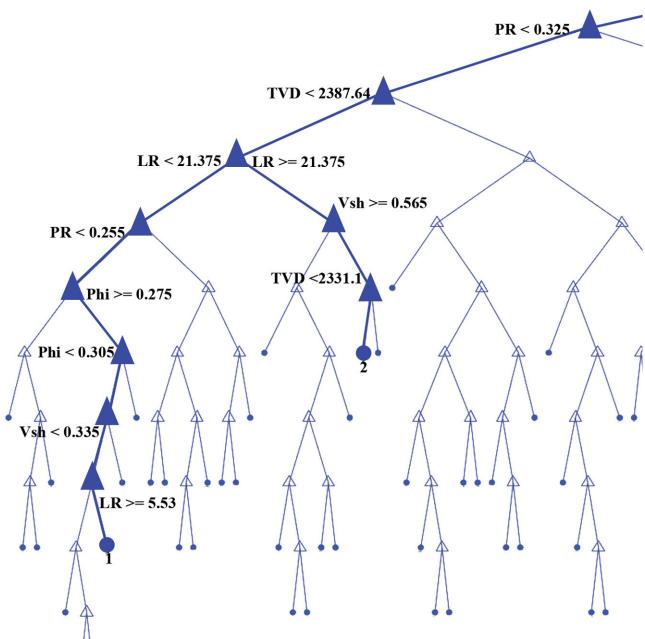


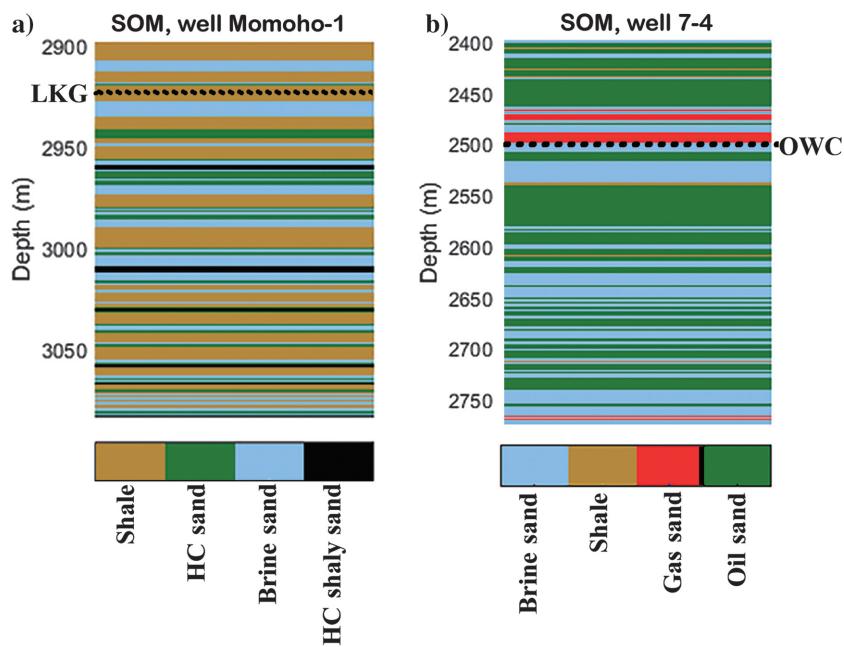
Figure 3. A part of one of the trees in a bagged-tree (BT) analysis trained on Heidrun data. Two observations leading to two example leaves are shown in bold. Leaves for this data set are (1) gas sand, (2) shale, (3) oil sand, and (4) brine sand. The specifying feature values for those observations are shown at the splitting nodes. TVD: true vertical depth, PR: Poisson's ratio, LR: lambda-rho, phi: porosity, V_{Sh} : shale volume.

Figure 2. Silhouette values for the Heidrun Field (a) and Kupe Field (b). The values are calculated for different cluster numbers as predicted by SOM networks. The higher value indicates the best separation between clusters and hence the optimum number of clusters for the clustering method. However, lower values indicate that more similarities exist between the separated clusters.

Table 1. Summary of BT and MLFN training results. The box plots in the accuracy range show the results of seven independent training sessions. The numbers on these plots are the minimum, average, and maximum, from bottom to top, and the line indicates the median. The top and bottom of the boxes are the third and first quartiles, respectively. The other rows show the results of the algorithm with the median accuracy. The accuracy is based on the cross-validated training error, and the success rates are the positive predictions for each class.

		Heidrun		Kupe	
No. of samples	1910		1986		
ML algorithm	MLFN	BT	MLFN	BT	
Accuracy range					
	90.4 89.7 88.8	96.5 95.82 95.4	87.9 85.22 82.9	89.4 88.78 88.2	
Accuracy (%)	89.9	95.7	85.5	88.7	
HC success rate (%)	95.4	97.5	86.2	89.8	
Non-HC success rate (%)	87.6	94.7	84.6	88.1	

Figure 4. The SOM-predicted lithofluid facies in example wells from Kupe Field (a) and Heidrun Field (b). The dashed lines show the lowest known gas (LKG) and oil-water contact (OWC) for the two wells.



tial effort to find the best nonANN algorithms. In this procedure, the data are subdivided into five sections (folds) during the training phase, and at each turn, one section is predicted by the rest. The accuracy of the algorithm is calculated based on the average of these cross-validation errors. In all of the trial sessions of training for this study, BT constantly had the highest accuracy among all the available ML algorithms.

Discussion and results

SOM clusters the data based on similarities without using labels (targets); thus, the results of the clustering need interpretive validations after being applied. Our interpretation of the implied lithofluid facies labels is based on investigating the results in crossplots and PDF histograms.

The best separation among SOM clusters occurred when the algorithm looked for four classes within each data set, according to the silhouette value. This was in agreement with our assumption about the target classes. However, a preliminary investigation of the SOM-driven lithofluid facies logs shows a disagreement between the predicted and actual fluid content (Figure 4). In most cases, the predicted HC column extends deeper than the lowest HC observed in the wells (OWC or LKG). In addition, in the Heidrun Field, gas sand only exists in well

6507/8-1, whereas thin layers of this class are predicted in other wells by SOM.

Ross and Cole (2017) compare the results of MLFN with unsupervised ANNs in a seismic facies mapping case and conclude that the latter is not well-suited for solving nonlinear problems as in facies classification. However, Zhao et al. (2015) argue that an unsupervised algorithm has the advantage of not being limited to a predefined number of classes, hence being a better choice in facies classifications in which unseen classes may occur in the data set. According to the goals of our study, the SOM results were not reliable; thus, it was not considered for simulating on the target wells.

Two supervised ML algorithms, BT and MLFN, were trained on two data sets. We trained each algorithm seven times independently (not retraining), and then we selected the median values to compare the accuracy of the two techniques. The reason behind multiple training sessions is to have a measure of the stability of each algorithm because ML algorithms usually (if not intentionally designed otherwise) make use of randomness at some point in training. For example, the initiation of weights in ANN, the observation order in BT, and sampling data to create validation subgroup(s) happen in a random state. This randomness causes the nonrepeatability of the exact training results in different runs of the algorithm. Accuracy values and ranges, and the success rate in predicting HC and non-HC units, are summarized in Table 1. The validation errors for MLFN and BT are reflected in their accuracy values.

An important thing to note in this table is the range or spread of accuracy results depicted by the heights of

the box plots. BT appears to be a more stable algorithm due to its lower variation in accuracies (smaller range) in both fields. It means that despite the inherent nonrepeatability of results each time that the algorithm is initiated, BT can be expected to have an almost predictable accuracy (approximately $\pm 0.5\%$). It also shows a higher overall and average accuracy in both fields. This capability of BTs in generalizing results with minimum overfitting lowers the risk of false discoveries.

The application of BTs can be viewed as going through a series of if-then rules that were initially extracted from the training data set. Thus, it is a natural choice to use BTs in wireline log interpretation because the log interpretation is a decision-making task based on how each log is behaving.

The success-rate values show the positive predictions within each category. For instance, in Kupe Field, 89.8% of HC units (two HC classes combined) predicted by BT belong to the HC units in the target log, which means that 10.2% of what is predicted as HC units in fact belong to nonHC units. Based on this parameter, both algorithms are more successful in predicting HC units than nonHC units. However, BT has higher success rates in predicting either HC or nonHC units in comparison with MLFN.

Based on these results, we chose the BT algorithm to predict the lithofluid facies logs in each field, the results of which are discussed in the following sections. Figure 5 shows the BT prediction of lithofluid facies logs in one well at each field, along with three wireline logs as reference. As shown in this figure and discussed

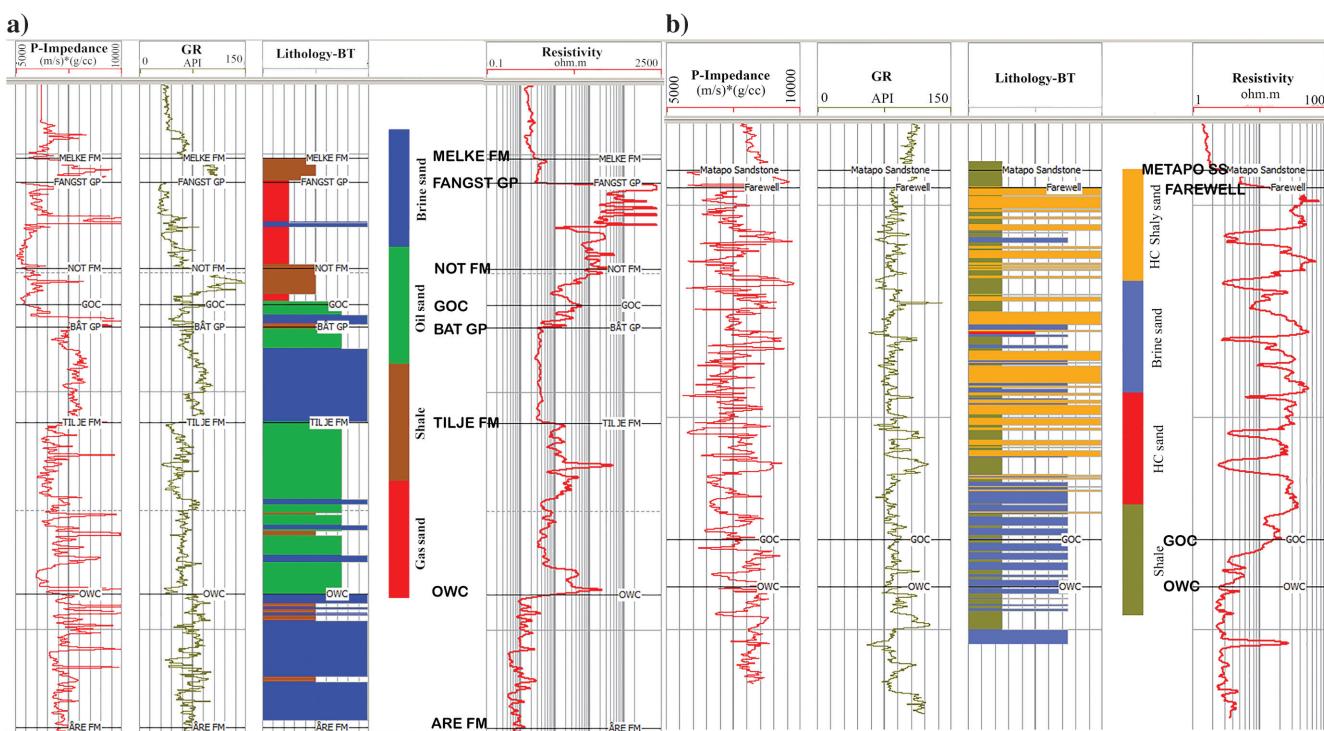


Figure 5. The BT prediction of the lithofluid facies for well 6507/8-1 of Heidrun Field (a) and well Kupe South-7 of Kupe Field (b). P-impedance, gamma ray, and deep resistivity logs are shown for reference.

before, there is a larger amount of shale content in the Kupe Field in comparison with the Heidrun Field. The shale content, either in the form of alternating layers or clay content within the sandstones, affects the wireline logs, which makes the qualitative log interpretations more difficult and less accurate in Kupe than in Heidrun. ML algorithms, due to their ability in handling nonlinearity in multiattribute problems, can increase the accuracy in such complicated cases.

Heidrun lithofluid facies logs

The data were analyzed qualitatively and using LMR and statistical analysis to find the well logs that can best predict the lithofluid facies. In all cases, PR, LR, and Z_p were the best features in distinguishing between all classes, whereas the other logs had varying performances in separating different classes. In all cases, shale and brine sand were found to be difficult to distinguish from each other. One reason could be that the shale class is underrepresented in the Heidrun data, leading to an incomplete characterization of the shale

class. An underrepresented class, however, does not necessarily have a lesser chance of receiving votes in BT because the algorithm uses the empirical probability of each class in training. The Fangst Group is mainly shallow-marine and fluvial sandstone deposited during the Late Jurassic regression in off-shore mid-Norway and includes one thin marine-shale layer in the Not Formation (Harris, 1989). Two shaly classes are included in data that are outside the reservoir: the overlying Melke marine-shale of the Viking Group and the underlying Ror and Tilje muddy sandstone and sandstone and shale from the Båt Group. However, these layers are not clean shales or thick enough to completely overcome the potential bias in the number of samples. A sandy-shale target sample, for example, is easy to be misclassified with brine sand due to their similar Z_p and LR values.

The BT algorithm was applied to wells 6507/7-3 and 6507/7-8 and the excluded section of 6507/7-4. The results are shown in Figure 6. The results demonstrate the capability of BT in predicting HC units: Predicted oil sands agree with the OWC depths, and there are no gas sands predicted falsely in the target wells. In addition, shale layers belonging to the Viking Group, Ror and Tilje Formations (Båt Group), and Not Formation, where they exist, are predicted correctly.

As shown in the confusion matrix (Figure 7a), the performance of the trained BT algorithm can be investigated in more detail. The success rate values shown in this matrix are averaged over cross validations, whereas one-fifth of the data were randomly selected and predicted by the rest of the data during a training session. The percentages are based on the predicted classes; a false discovery shows the percentage of a predicted class that does not belong to that class in the target log. For example, the false discovery of gas sand shows that 4% of the predicted gas sands were, in fact, shale (1%), and brine sand (3%). Oil sands also were correctly classified in 98% of the predictions, with only 1% and less than 1% of the predictions belonging to brine sands and shale, respectively. The second and fourth columns show that most of the false discoveries occurred between shale and brine sands.

Because in this field we have access to the target lithofluid facies logs, we expanded these calculations to our predictions of the target wells. The confusion matrix in Figure 7b shows the positive predictions versus the false discoveries. In addition, true classification and misclassification rates are shown in the last column and the last row of the confusion matrix, respectively. A misclassification rate is the percentage of a true class that is predicted as other classes. The positive prediction rates confirm that predicted HC units by BT belong to the right classes with high positive prediction rates: 92.1% for oil sands, and 0% for gas sands. Note that the target wells do not have any gas-sand layers and BT has not falsely discovered any other classes as gas sands in these wells.

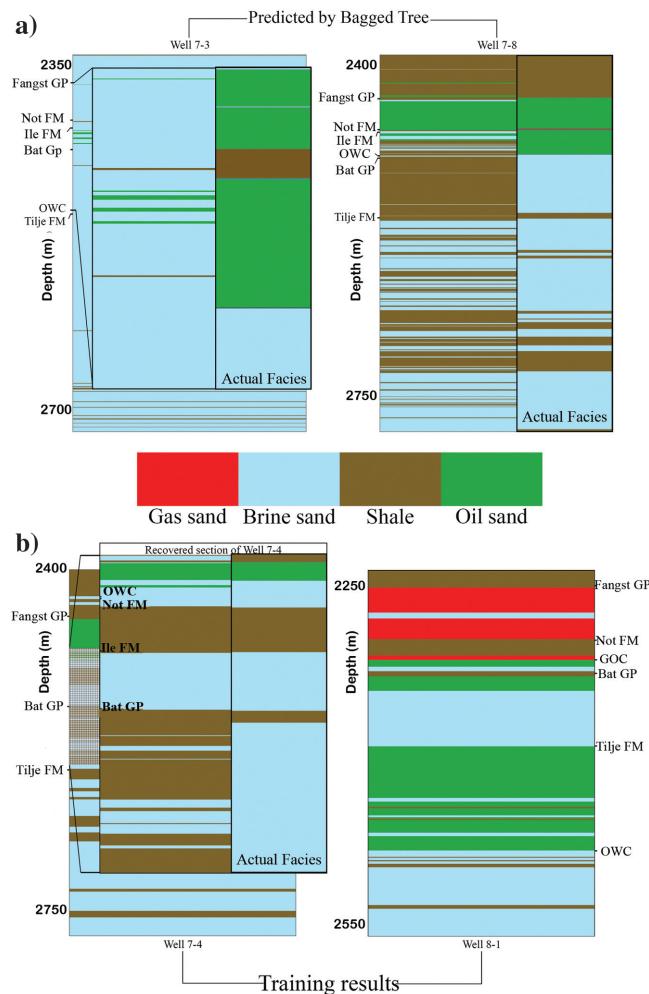


Figure 6. (a) The BT-predicted logs (6507/7-3 and 7-8) and (b) the predicted section of 6507/7-4 (left, magnified). The rest of 6507/7-4 and well 6507/8-1 were used to train the BT.

Similar to the training results, the shale and brine-sand classes account for most of the false discoveries between each other (see the second and fourth row on the brine sand and shale target columns, respectively). The false discoveries between shale and brine sand can be explained by the previously discussed issue of the similarities between the shale and brine sand in this data set.

However, the overall low-accuracy rate in this model (63.2%) is not only due to the shale/brine-sand false discoveries. The last row of this matrix shows the misclassification rate in the red (bottom) numbers. The highest misclassification rate belongs to the oil sands indicating that 60% of actual oil sands are misclassified as brine sand, and 9% as shale. This is mostly caused by well 6507/7-3 because taking it out of the calculations drops the misclassification of oil sands from 60% to 25.2% (Figure 7c). The chart in Figure 8 shows the role of each target well in the overall accuracy rate of BT. Because the errors are normalized by the number of samples at each well, the stacked bars confirm that most of the error in the oil-sand class is caused by well 6507/7-3.

The reason behind the distinct behavior of this well is that, in creating the lithofluid facies logs, we relied on drilling reports to validate the fluid content. Consecutively, a thick layer of oil sands (approximately 100 m) is designated for an interval with HC shows as described in the reports. However, the wireline logs do not show such a consistent and significant anomaly over the reservoir interval. For example, formation resistivity logs (RTF) available only in wells 6507/7-4,

6507/7-A-53, and 6507/7-A-17 show this discrepancy among these wells and well 6507/7-3 (Figure 9). The RTF anomaly of the Fangst reservoir in this well is not as high or as blocky as observed in other wells. We can argue that the BT prediction, in contrast to the reference target, is correctly indicating a low-quality reservoir section encountered by this well, which is in agreement with the results of an ANN-driven PR model of this field (Keynejad et al., 2017).

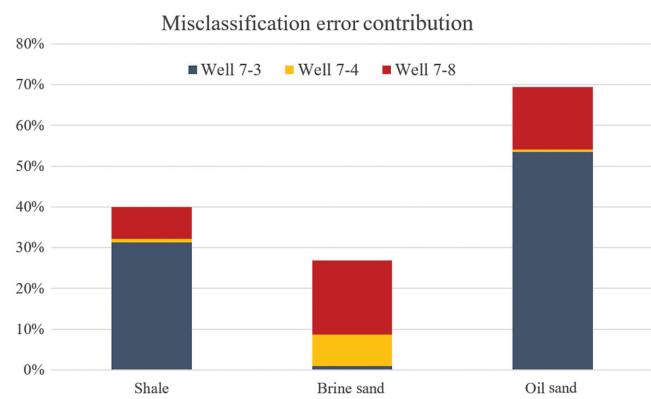


Figure 8. Error contributed by each target well (6507/7-3, 6507/7-4, and 6507/7-8) in BT's overall prediction error. The overall misclassification errors are 40%, 26.9%, and 69.5% for the shale, brine-sand, and oil-sand classes, respectively. Note that most of the misclassified oil sands are caused by data from well 7-3.

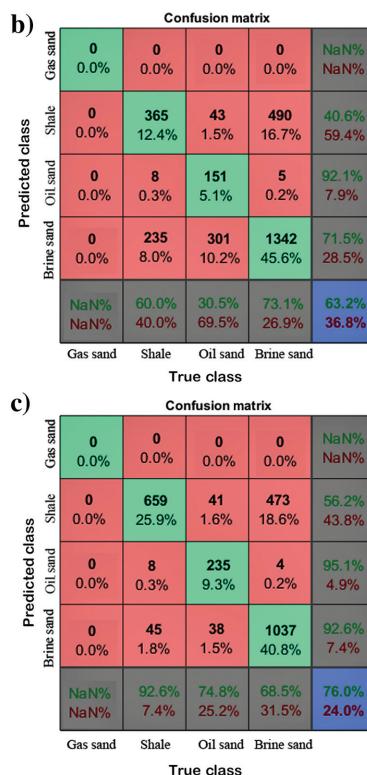
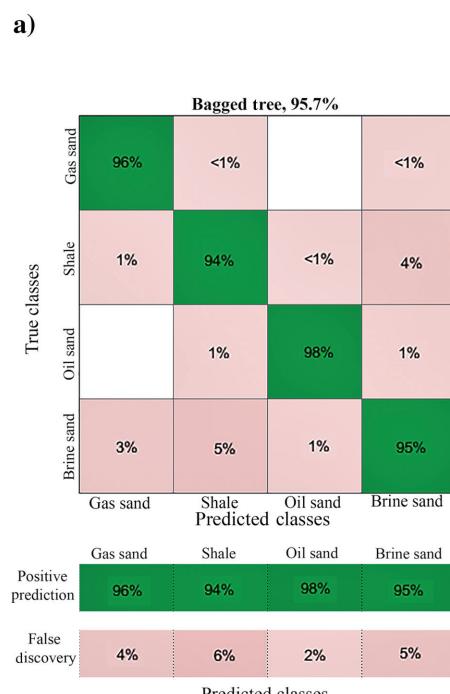


Figure 7. (a) Confusion matrix of trained and cross-validated BT results on Heidrun data with an average accuracy of 95.7%. On the right (b) the prediction results are shown for all target wells and (c) all except well 6507/7-3. The cells in the dark columns in (b and c) include the positive prediction in green (top) numbers and the false discovery in red (bottom) numbers. The cells in the dark rows in (b and c) show the correct classification in green (top) numbers and the misclassification in red (bottom) numbers. The blue boxes in (b and c) show the overall accuracy and error in green (top) and red (bottom) numbers, respectively.

Kupe lithofluid facies logs

The confusion matrix in Figure 10 shows a high positive predictive rate for HC-bearing clean sands (92%). The predicted shaly sands with HC-shows have a lower rate with 77% of the predictions being correctly classified. The internal percentages of this class (the fourth column) show that of the 23% false discoveries, 13% belonging to the shale class. This was predictable when this class was defined to be an indicator of low-quality reservoir sand, with features between a clean reservoir sand and shale classes in the target log. Note also that most of the clean sand false discoveries are brine sand rather than HC shaly sand. This suggests that the properties of the HC shaly sand are farther from the clean HC sand than for the brine sand.

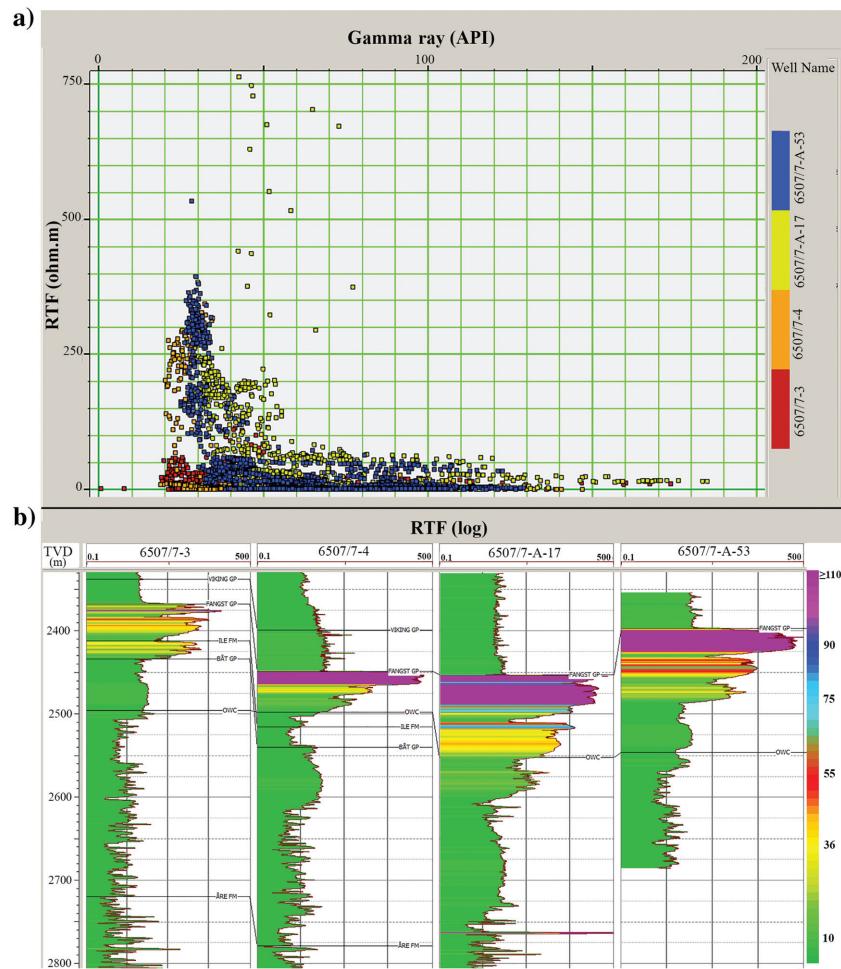
The predicted lithofluid facies logs for four target wells are shown in Figure 11. To qualitatively compare the results, the available interpreted composite log and the HC shows are also included. As seen in this figure, there are other interpreted lithofacies that we have not included in our study because they are neither related to our goal, nor are they large enough to affect the algorithm's performance significantly.

The comparison suggests the following points:

- 1) Predicted HC units are in good agreement with HC shows or interpreted fluid contacts; no HC unit is predicted below the expected reservoir zone.
- 2) Predicted shales, especially the sealing shale overlying the Farewell reservoir, are consistent with the interpreted layers of claystone and siltstone and/or are proportional to the thickness or amount.
- 3) The HC shaly sands predicted in well Kupe South 7 are consistent with the interpreted argillaceous sandstones.

Kupe South 5 is reported to have HCs shows over an interval of approximately 36 m, which is not reflected in the BT predictions. The interpreted lithofacies column from drilling reports shows that alternating shale layers comprise half of the Farewell Formation in this well, the rest of which are carbonaceous sandstones. Such discrepancies should be investigated through core sample analysis to better understand whether this is only a misclassification. The drilling documents of this well do not mark the Farewell sandstones as a reservoir, and they report that severe calcite cementations and formation

Figure 9. (a) Crossplot of the true formation resistivity (RTF, ohm-m) versus gamma ray (API). (b) The RTF logs plotted for the same well, colored by the RTF amplitude at the reservoir depth interval. The RTF curves are shown in logarithmic scale. The color scale for the curves is cropped at 110 ohm-m to better depict the variations at the reservoir interval (i.e., RTF values of 110–500 ohm-m all are in purple).



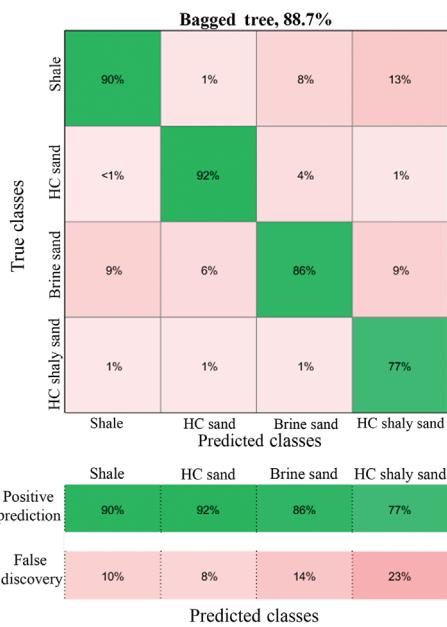


Figure 10. Confusion matrix, showing the cross-validated results of training BT algorithm on Kupe Field data, with 88.7% average accuracy.

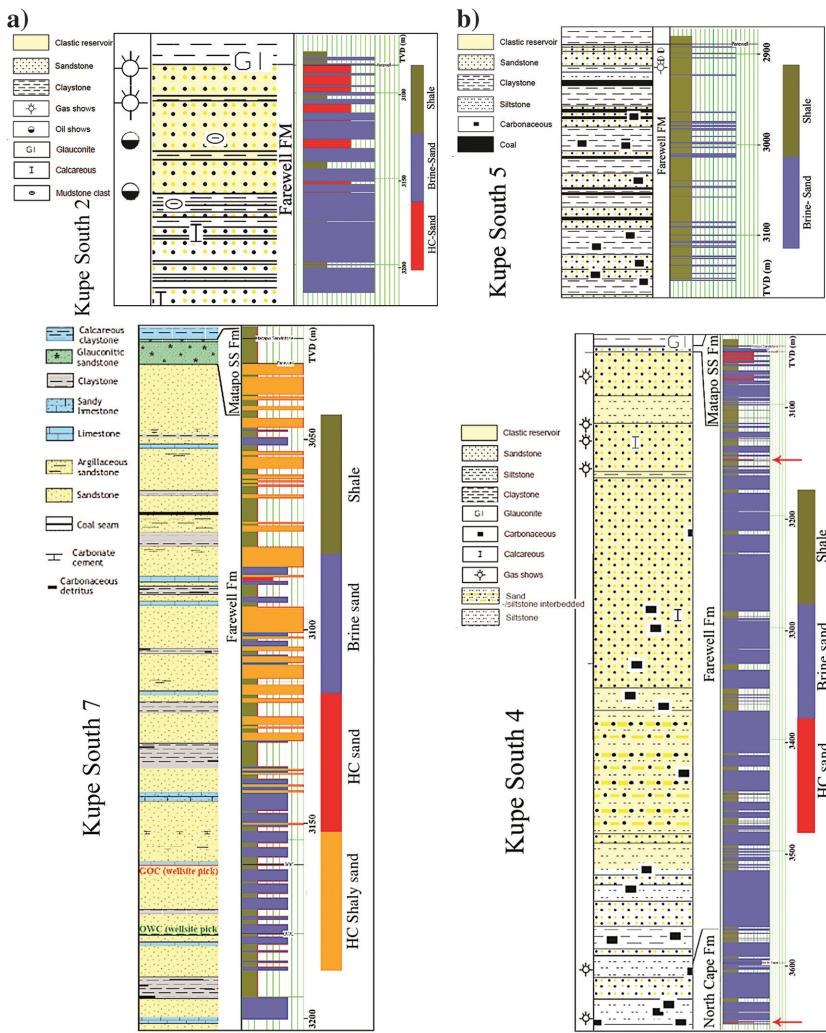


Figure 11. The BT-predicted lithofluid facies logs in Kupe Field (b), compared with the interpreted lithofacies (a). Note that the BT-predicted results provide more resolution compared to the qualitatively interpreted columns. Red arrows for Kupe South 4 point to very thin HC sands predicted by BT.

of authigenic clay minerals have affected the reservoir quality (New Zealand Petroleum and Minerals, 2017).

Conclusion

Our study suggests the usage of ML algorithms along with the expert's knowledge to gain objective insight of the subsurface properties while tuning the algorithm and resultant interpretations based on the expectations and information about the reservoir. The main advantage of ML is its power in resolving complex, multicriteria problems. Another advantage of this approach is that, once trained, it can be applied to wells with minimum available logs because the target logs used in training are designed based on reliable information and a complete package of logs in the training wells. We have compared an unsupervised and two supervised approaches and applied the most successful algorithm on two HC fields to interpret its results.

In each field, we used different wireline logs to create lithofluid facies logs for training wells. The facies defined in this study are not detailed stratigraphic classes; instead, the facies were placed into simplified, inclusive groups with the main aim of distinguishing HC-bearing units from the rest, in accordance with

the reservoir's properties. Although this approach can be of more importance in the appraisal and development phases, ML applications with more class variations based on the stratigraphy and/or depositional settings can be similarly designed for explorational stages. Note that misclassifications are inevitable in any sort of indirect assessment of the facies because none of them are completely distinct from the others, physically and/or compositionally.

Although the unsupervised approach (SOM) was helpful as a preprocessing step to establish the idea about the number of clusters and the predictive attributes without the need to know the targets, the results do not seem to be reliable enough. This approach, coupled with knowledge-driven techniques, can especially be of help in cases in which there are no cored wells near the study area.

Both supervised methods could successfully predict HC units in the cross-validation training phase in both fields, with BT having a higher overall and HC-related accuracy rate. We expect ANNs to be more successful in problems in which the relationship among attributes and targets is more complicated than being mapped with a system of if-then rules. Using seismic attributes for mapping seismic facies is an example in which the rules relating the features to the target, though basically understood, cannot be directly drawn at any given sample.

A more detailed study of the prediction results for trained BTs showed a very low false discovery rate for oil sands (7.9%) and no false discovery for gas sands in Heidrun Field. However, about two-thirds of the existing oil sands were misclassified as brine sands, with well 6507/7-3 as the major contributor to this error. The prediction, though deviating from the previously interpreted fluid column (hence causing the error), is in agreement with the low formation resistivity values and the relatively high Poisson's ratio at this well location.

Qualitative investigations of Kupe Field results also demonstrate no false discoveries for HC units. The HC-shaly sand class was defined for low-quality reservoir sands and was expected to be mostly misclassified either as clean HC sands or shale. It was successfully predicted at Kupe South 7, which is a deviated well from the training well that contained this class. Thin layers of predicted different classes are in accordance with variations in the associated index logs, which can mean a more reliable resolution than the qualitative interpretations. The HC shows reported in well Kupe South 5 that occur in a very low-porosity section of the Farewell Formation, which was not marked as reservoir, were not predicted by BT. For such reservoir conditions, predicting an HC unit would more likely be a falsely discovered reservoir and cause more harm in field development plans if decisions are based on misleading results.

The apparent misclassification of HC units as nonHC units in this study can be interpreted using core samples and other comprehensive field studies. If the misclassifications are real, they can cause an underestimation of

the reservoir capacity. In general, to lower any misclassification in predicting lithofluid facies in a reservoir, the observed cases of misclassification should be interpreted and addressed during the training phase.

The advantage of using ML algorithms is their power in scrutinizing large data sets with multidimensional, complicated, interrelated attributes. This study showed that ML algorithms can be used in classifying lithofluid facies to predict HC-bearing units, with minimal to no risk of HC false discoveries. The proposed approach of using ML algorithms, trained and tuned based on an expert's knowledge about the reservoir, can be modified and applied on future wells in a HC field to distinguish the pay zone.

Acknowledgments

The authors thank MathWorks for providing MATLAB, Simulink under the academic license to the University of Arizona. We also thank CGG Services (U.S.) Inc. for providing licenses for their GeoSoftware package, including the HampsonRussell suite of modeling and analysis software. We gratefully acknowledge ConocoPhillips and its partners, Equinor, Petoro, and Eni for providing the Heidrun Field data set used in this research, with a special thanks to G. Hill for his support and assistance in this procedure. We are grateful to A. Hughes for providing us with the Kupe Field database, originally issued by New Zealand Petroleum & Minerals.

Data and materials availability

Data associated with this research are confidential and cannot be released.

References

- Aleardi, M., and F. Ciabbari, 2017, Application of different classification methods for litho-fluid facies prediction: A case study from the offshore Nile Delta: *Journal of Geophysics and Engineering*, **14**, 1087–1102, doi: [10.1088/1742-2140/aa7301](https://doi.org/10.1088/1742-2140/aa7301).
- Al Moqbel, A., and Y. Wang, 2011, Carbonate reservoir characterization with lithofacies clustering and porosity prediction: *Journal of Geophysics and Engineering*, **8**, 592–598, doi: [10.1088/1742-2132/8/4/011](https://doi.org/10.1088/1742-2132/8/4/011).
- Al-Mudhafar, W., 2017, Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms: *Journal of Petroleum Exploration and Production Technology*, **7**, 1023–1033, doi: [10.1007/s13202-017-0360-0](https://doi.org/10.1007/s13202-017-0360-0).
- Breiman, L., 1996, Bagging predictors: *Machine learning*: Springer **24**, 123–140.
- Breiman, L., 2001, Random forests: *Machine learning*: Kluwer Academic Publishers **45**, 5–32.
- Chang, H., D. C. Kopaska-Merkel, H. Chen, and S. R. Durrans, 2000, Lithofacies identification using multiple adaptive resonance theory neural networks and group decision expert system: *Computers and Geosciences*, **26**, 591–601, doi: [10.1016/S0098-3004\(00\)00010-8](https://doi.org/10.1016/S0098-3004(00)00010-8).

- Close, D., R. Taylor, and S. Nixon, 2016, A case study in quantitative interpretation ambiguity, lambda-mu-rho, and rock-physics modeling in the Otway Basin, Australia: *The Leading Edge*, **35**, 43–50, doi: [10.1190/tle35010043.1](https://doi.org/10.1190/tle35010043.1).
- Goodway, B., T. Chen, and J. Downtown, 1997, Improved AVO fluid detection and lithology discrimination using lamé petrophysical parameters; “ $\lambda\rho$ ”, “ $\mu\rho$ ”, & “ λ/μ Fluid Stack”, from P and S inversions: 67th Annual International Meeting, SEG, Expanded Abstracts, 183–186, doi: [10.1190/1.1885795](https://doi.org/10.1190/1.1885795).
- Harris, N. B., 1989, Reservoir geology of Fangst Group (Middle Jurassic), Heidrun field, offshore Mid-Norway: *AAPG Bulletin*, **73**, 1415–1435.
- Keynejad, S., M. Sbar, and R. Johnson, 2017, Comparison of model-based generalized regression neural network and prestack inversion in predicting Poisson's ratio in Heidrun Field, North Sea: *The Leading Edge*, **36**, 938–946, doi: [10.1190/tle36110938.1](https://doi.org/10.1190/tle36110938.1).
- Kohonen, T., 1987, Self-organization and associative memory: Springer-Verlag.
- Liu, R., B. Zhang, and X. Wang, 2017, Patterns classification in assisting seismic-facies analysis: 87th Annual International Meeting, SEG, Expanded Abstracts, 2127–2131, doi: [10.1190/segam2017-17795000.1](https://doi.org/10.1190/segam2017-17795000.1).
- Martin, K. R., J. C. Baker, P. J. Hamilton, and G. P. Thrasher, 1994, Diagenesis and reservoir quality of paleocene sandstones in the Kupe South field, Taranaki basin, New Zealand: *AAPG Bulletin*, **78**, 624–643.
- Morton, A., C. Hallsworth, D. Strogen, A. Whitham, and M. Fanning, 2009, Evolution of provenance in the NE Atlantic Rift: The early-middle Jurassic succession in the Heidrun field, Halten Terrace, offshore Mid-Norway: *Marine and Petroleum Geology*, **26**, 1100–1117, doi: [10.1016/j.marpetgeo.2008.07.006](https://doi.org/10.1016/j.marpetgeo.2008.07.006).
- New Zealand Petroleum and Minerals, 2017, New Zealand Petroleum Exploration Data Pack, New Zealand Petroleum & Minerals, Norwegian Petroleum Directorate, Heidrun Field, <http://factpages.npd.no/factpages>, accessed September 2017.
- Pang, J., and J. Collen, 1996, Well log analysis of the Kupe South field, Taranaki basin, New Zealand: SPE Asia Pacific Oil and Gas Conference.
- Poulton, M. M., 2002, Neural networks as an intelligence amplification tool: A review of applications: *Geophysics*, **67**, 979–993, doi: [10.1190/1.1484539](https://doi.org/10.1190/1.1484539).
- Qi, L., and T. R. Carr, 2006, Neural network prediction of carbonate lithofacies from well logs, big bow and Sand Arroyo Creek fields, Southwest Kansas: *Computers and Geosciences*, **32**, 947–964, doi: [10.1016/j.cageo.2005.10.020](https://doi.org/10.1016/j.cageo.2005.10.020).
- Ross, C. P., and D. M. Cole, 2017, A comparison of popular neural network facies-classification schemes: *The Leading Edge*, **36**, 340–349, doi: [10.1190/tle36040340.1](https://doi.org/10.1190/tle36040340.1).
- Rousseeuw, P. J., 1987, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis: *Journal of Computational and Applied Mathematics*, **20**, 53–65, doi: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Torres, A., and J. Reverón, 2014, Integration of rock physics, seismic inversion, and support vector machines for reservoir characterization in the Orinoco oil belt, Venezuela: *The Leading Edge*, **33**, 774–782, doi: [10.1190/tle33070774.1](https://doi.org/10.1190/tle33070774.1).
- Young, K. T., and R. H. Tatham, 2007, Lambda-mu-rho inversion as a fluid and lithology discriminator in the Columbus basin, Offshore Trinidad: 77th Annual International Meeting, SEG, Expanded Abstracts, 214–218, doi: [10.1190/1.2792413](https://doi.org/10.1190/1.2792413).
- Zhao, T., V. Jayaram, K. J. Marfurt, and H. Zhou, 2014, Lithofacies classification in Barnett shale using proximal support vector machines: 84th Annual International Meeting, SEG, Expanded Abstracts, 1491–1495, doi: [10.1190/segam2014-1210.1](https://doi.org/10.1190/segam2014-1210.1).
- Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: *Interpretation*, **3**, no. 4, SAE29–SAE58, doi: [10.1190/INT-2015-0044.1](https://doi.org/10.1190/INT-2015-0044.1).
- Zhao, T., F. Li, and K. Marfurt, 2017, Constraining self-organizing map facies analysis with stratigraphy: An approach to increase the credibility in automatic seismic facies classification: *Interpretation*, **5**, no. 2, T163–T171, doi: [10.1190/INT-2016-0132.1](https://doi.org/10.1190/INT-2016-0132.1).

Biographies and photographs of the authors are not available.