Preemptive Detection of High Water-Cut Wells in Delaware Basin using a Joint Unsupervised and Supervised Learning Approach

Jonathan Foster[1,1], Siddharth Misra[1,1], YUSUF FALOLA[1,1], and Mukul Bhatia[1,1]

[1]Texas A&M University

November 30, 2022

Abstract

High water cut has been an issue in the Delaware basin for many years now. Volume of produced water continue to increase, resulting in adverse environmental impacts and higher reservoir-management costs. To address these problems, a data-driven workflow has been developed to pre-emptively identify the high water-cut wells. The workflow uses unsupervised pseudo-rock typing followed by supervised classification trained on well logs from 17 wells in the Delaware basin. The workflow requires a suite of 5 well logs from a 500-ft depth interval surrounding the kick-off points of these wells, which includes 200 ft above and 300 ft below the KOP. First, the well logs are clustered into 5 pseudo-rock types using multi-level clustering. Using statistical features extracted from these 5 pseudo-rock types, 3 supervised classifiers, namely K-nearest neighbor, support vector machine, and logistic regression, are trained and tested to detect the high water-cut wells. Over 100 cross validations, the 3 classifiers perform at a median Matthew's Correlation Coefficient (MCC) of 0.90. The kurtosis of the neutron porosity log response of the pseudo-rock type A0, interpreted as a shale lithology, is the most The submitted paper is currently under review. Dr. Sid Misra is the lead investigator on this topic. informative/relevant signature associated with high water cut. Next, the presence of pseudo-rock type A1, interpreted as high-permeability lithology, is an informative signature of low water-cut wells. The kurtosis of the density porosity log response of the pseudo-rock type B0, interpreted as carbonate lithology, and the presence of pseudo-rock type B1, interpreted as a tight sandstone lithology, serve as informative signatures for differentiating high water cut wells from low water cut wells.

Preemptive Detection of High Water-Cut Wells in Delaware Basin using a Joint Unsupervised and Supervised Learning Approach

Jonathan Foster, Siddharth Misra^*, Yusuf Falola^, Mukul Bhatia*+*

^Harold Vance Department of Petroleum Engineering, Texas A&M University, USA

*The Department of Geology and Geophysics, Texas A&M University, USA

+Berg-Hughes Center for Petroleum & Sedimentary Systems, Texas A&M University, USA

Corresponding author: misra@tamu.edu

Abstract

High water cut has been an issue in the Delaware basin for many years now. Volume of produced water continue to increase, resulting in adverse environmental impacts and higher reservoir-management costs. To address these problems, a data-driven workflow has been developed to pre-emptively identify the high water-cut wells. The workflow uses unsupervised pseudo-rock typing followed by supervised classification trained on well logs from 17 wells in the Delaware basin. The workflow requires a suite of 5 well logs from a 500-ft depth interval surrounding the kick-off points of these wells, which includes 200 ft above and 300 ft below the KOP. First, the well logs are clustered into 5 pseudo-rock types using multi-level clustering. Using statistical features extracted from these 5 pseudo-rock types, 3 supervised classifiers, namely K-nearest neighbor, support vector machine, and logistic regression, are trained and tested to detect the high water-cut wells. Over 100 cross validations, the 3 classifiers perform at a median Matthew's Correlation Coefficient (MCC) of 0.90. The kurtosis of the neutron porosity log response of the pseudo-rock type A0, interpreted as a shale lithology, is the most

The submitted paper is currently under review. Dr. Sid Misra is the lead investigator on this topic.

informative/relevant signature associated with high water cut. Next, the presence of pseudo-rock type A1, interpreted as high-permeability lithology, is an informative signature of low water-cut wells. The kurtosis of the density porosity log response of the pseudo-rock type B0, interpreted as carbonate lithology, and the presence of pseudo-rock type B1, interpreted as a tight sandstone lithology, serve as informative signatures for differentiating high water cut wells from low water cut wells.

1 INTRODUCTION

1.1 Background

The widespread implementation of hydraulic fracturing and lateral drilling has led to vast expansion of drilling in the Permian basin. This boom in oil and gas production was unexpectedly followed by high water production. The percentage of water produced from a hydrocarbon well is commonly referred to as a "water-cut." A study found that out of 10,000 shale-oil, unconventional wells in the Permian basin, a quarter of them have water-cuts as high as 70% [1]. This is particularly an issue in the western-most portion of the Permian basin, the Delaware basin. In the Delaware, water-to-oil ratios reach as high as 10:1 [2]. This produced water has become a huge logistical issue for operators in the region as water management fees continue to rise, with IHS Markit estimating a cost of $12.2 billion to Permian operators in 2018.

1.1.1 Motivation

The unit cost rise for a barrel of produced water is over $5.00/bbl [2]. If cost increases continue as projected, approximately 20% of all unproduced barrels of oil in the Permian will become non-commercial. Reusing the produced formation water has the potential to offset some of the

water management costs, but in exceptionally high water-cut wells operators are unable to reinject all produced volumes at a low cost [2]. High water-cut wells produce excess water, containing oil residues, sand or mud, naturally occurring radioactive materials, fracking chemicals, salts, and organic compounds. There are environmental concerns surrounding the produced water from the Permian. These produced waters have failed chemical examinations to determine safety levels for drinking and irrigation [3]. The reinjection of produced water into the subsurface through saltwater disposal (SWD) wells is a common water-management approach [4]. Some propose to inject these fluids into the deeper Ellenberger formation; however, there is a danger of inducing seismic activity in the region [4]. High water-cut wells are susceptible to adverse conditions, such as corrosion, scaling, salt deposition, asphaltene/wax deposition, and culminating with the need for safe water storage, management and disposal. Excess water production from oil wells in shale plays has large-scale adverse environmental consequences. There is no fundamental understanding & scientific consensus on the factors and processes influencing the water cuts of wells drilled in the Delaware Basin and several other U.S. shale plays.

1.1.2 Objective

In order to help mitigate issues of water management and potential environmental hazards, a data-driven workflow has been designed with the intention of preemptively detecting the high water-cut wells before production. Should an operator have reliable insight into a well's potential water-cut, they will be able to make an informed decision on the economic viability of the given well and better determine what is in their best interest. In addition to the data-driven workflow, we investigated the geologic factors which may be contributing to high water-cuts in the Delaware basin. As an extension from previous work [5], this data-driven workflow utilizes unsupervised learning for the purposes of predicting pseudo-rock types from well log data. These

pseudo-rock types significantly improve the robustness and geological consistency of the subsequent detection of high water-cut wells using supervised learning.

1.2 Well Log Suite

This data-driven workflow uses depth-based well logs to train unsupervised learning methods to predict the pseudo-rock types. Following that, statistical features extracted from well logs for each pseudo-rock type are used in the supervised learning for relating these statistical features with the target derived from production data. The well log suite which has shown reasonable success is the conventional “triple combo” well log suite, which is available in several wells drilled in the Permian basin. The triple combo log suite consists: Gamma ray (GR), porosity, and resistivity. The porosity logs used in the data-driven workflow are density porosity (DPHI) and neutron porosity (NPHI). The resistivity logs are shallow resistivity (ILS) and deep resistivity (ILD). This well log suite is popular due to the fact you can gain relatively deep insight into the rock properties of your target using a small amount of well logging tools. Gamma ray provides lithology information based on radioactive material within the rock body, which is closely associated with clay content and shale volumes. The porosity logs provide information on the storage available for subsurface fluids to saturate the pores of the target rock bodies. Neutron porosity can serve as an indicator of clay-bound water. The resistivity logs detail the fluid saturation of the rock body based on the inverse of its conductivity. The separation between shallow and deep resistivity logs serve as an indicator of the permeability of the formation.

1.3 Production Data and Produced Water Volume

The production data from the horizontal wells in the Delaware basin data set is used to generate the target, which is a categorical variable based on the relative volume of water production with respect to total volume of fluids produced from a well. Such a target is essential for the supervised learning required to relate the well-log based, pseudo-rock type specific statistical features to the target. For the desired task of detecting high water-cut wells, the target is either a high water producer (HWP) or low water producer (LWP). Due to the fact the wells are located in Texas, the water produced from each well is not reported. To remedy this lack of information, the water volumes were calculated based on the number of barrels of water produced for each well during well tests. Well tests within a lease are averaged and then approximated to a synthesized volume of water for the lease. This lease-sized volume of water is allocated to each well based on its calculated water-cut percentage. Although it is not reported water volume, this method of calculation has been tested on a blind spot-check basis and has proven to be very accurate.

1.4 Target Label Definition

From the calculated water volumes, 4 different time intervals were used to calculate the water production ratio (WPR). These 4 different time intervals are: 2 months, 6 months, 1 year, and 2 years of production from the beginning of a well's production. The amount of oil, gas, and calculated produced water was provided for each of these time intervals. The production values from these 4 separate time intervals were averaged, then using the following formula (*eq. 1*) the WPR is calculated for each well:

$$WPR = \frac{Produced\ Water}{Total\ Produced\ Fluid} \quad \text{(eq. 1)}$$

Once the WPR is calculated, thresholds were used to define two primary well classes as the target: High water producer (HWP) and low water producer (LWP). A HWP is defined as any well with a WPR greater than or equal to 0.70 and a LWP is defined as any well with a WPR lower than 0.50.

2 METHODOLOGY

2.1 Generalized Workflow

The generalized proposed workflow is presented in *Fig. 1*. Our proposed workflow consists of two primary portions: the target definition and the feature definition. The target definition is discussed in section *1.4.* The feature section of the workflow is split into the following components: well log extraction, log transformation, outlier detection, multilayer clustering, and feature extraction. This workflow was designed to use unsupervised learning to automatically cluster the well log samples, which are sampled at every ½ foot typically, into separate pseudo-rock type. The purpose of the clustering procedure was to determine which rock properties best described and separated the lithologies within our interval of interest. Once the lithologies were determined, features were extracted with respect to each pseudo-roc type and some additional features based on depth. These features are designed to describe petrophysical properties on a per pseudo-rock type per well basis. These extracted features were then used to train supervised learning methods to classify and predict whether or not a given well will be a HWP or an LWP.

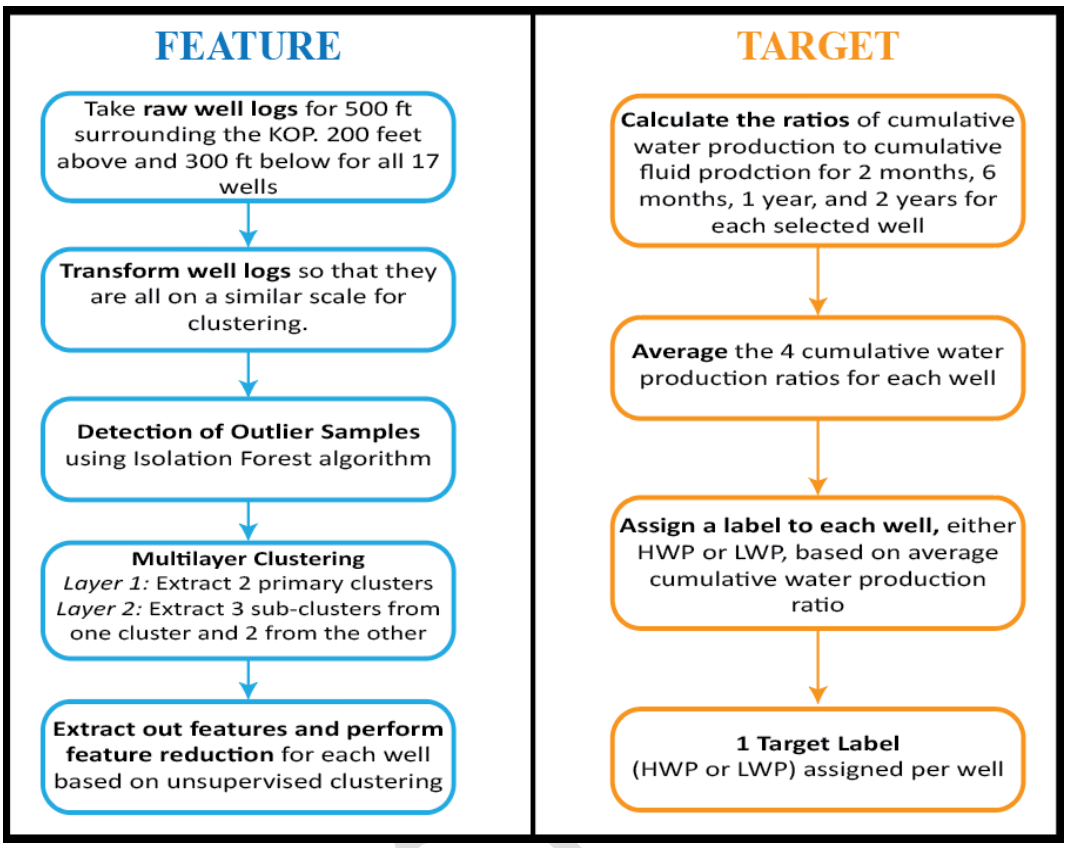

Fig. 1: Generalized data-driven workflow designed to utilize unsupervised learning in combination with supervised learning to preemptively detect the high water-cut wells.

2.2 Kick-Off Point Calculation

The interval of interest within the logged wellbore section of the wells in our data was determined by trial and error. By testing the predictability of features extracted out from well logs from various regions of the wellbore, it has been determined that the depth intervals nearest to the kick-off point (KOP) are the most effective descriptors for predicting relative water production. The typical definition for the KOP, and the definition used in this analysis, is the point at which a vertical wellbore begins transitioning into the horizontal wellbore. To automate this process and reduce user bias, the KOP was calculated algorithmically for each well. Using

trajectory data available, the inclination of the drill-bit was used as a guide to determine how horizontal the wellbore was at any depth. The depth of the KOP must be calculated in measured depth (MD), as opposed to true vertical depth, due to the fact well logs are presented in MD as well. The method of calculation used counts the number of depth samples in a row where the inclination is greater than 20 degrees, where 0 degrees is vertical. There are many variations in wellbore type in the Delaware basin, however this method has proven to be reliable in KOP calculation.

2.3 Interval of Interest

In our previous study [5], it was determined that the most informative features extracted from well logs in differentiating HWPs and LWPs were taken from the region nearest to the KOP in a given wellbore. Building off of these results, the well log data chosen to prioritize in this data-driven workflow also surrounds the KOP. The interval of interest of a given wellbore consists of 500 ft of well log data, using 200 ft of well log data above the KOP and 300 ft below the KOP. As well log readings are taken at every ½ foot interval, this depth interval provided us with 1000 samples of rock data from each well. Requiring this exact region of the wellbore to have well log data available resulted in the data set of 17 wells which were utilized in this analysis. Both intervals of interest are illustrated in *Fig. 2*.

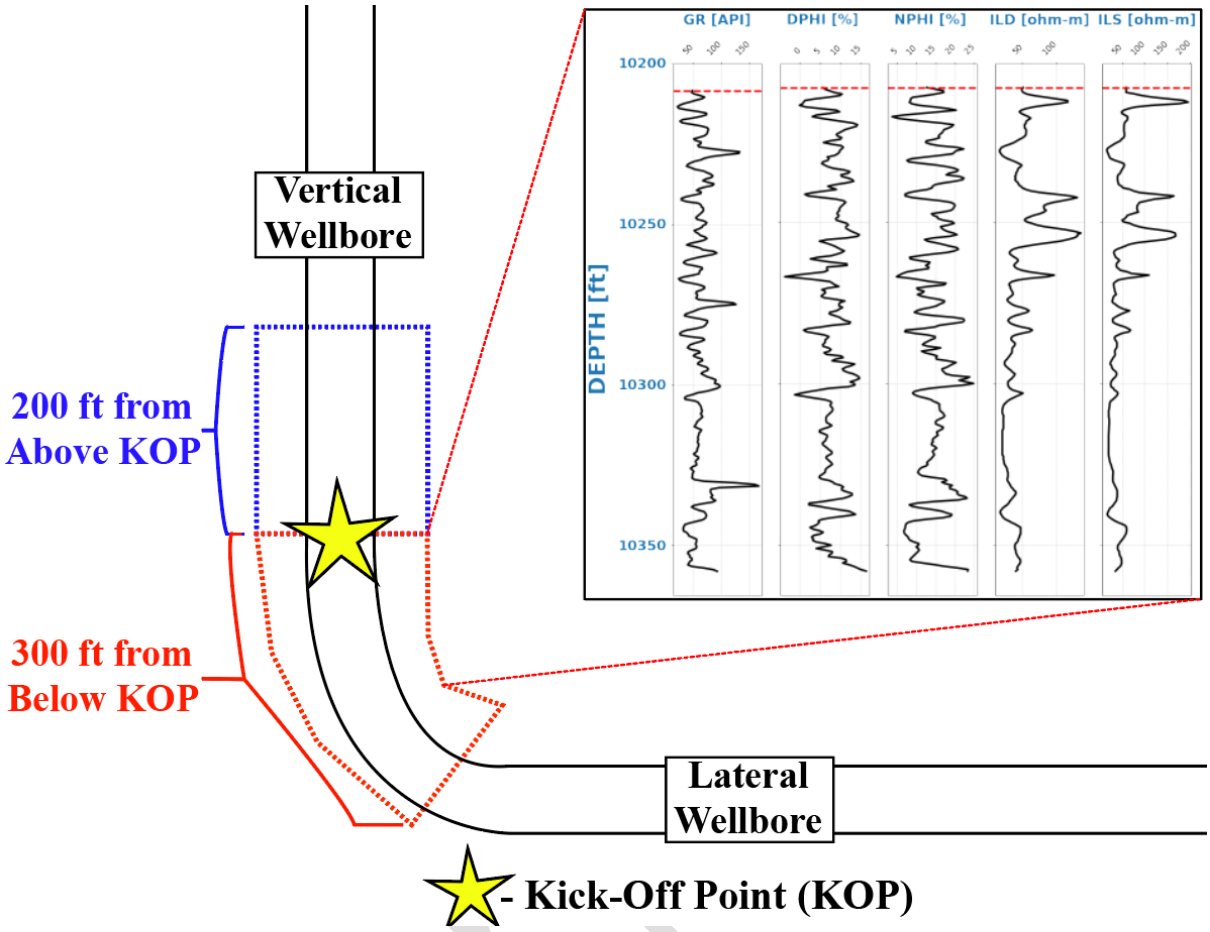

Fig. 2: Illustration of region(s) from where well log data was extracted. 200 ft from above kick-off point and 300 ft below the kick-off point were used in this data-driven workflow.

2.4 Data Processing

2.4.1 Scaling and Transforming Data

As with any machine learning workflow, it is optimal to perform feature engineering steps to construct the data into something resembling a Gaussian distribution centered at 0. This is necessary due to underlying assumption of a Gaussian distribution which most machine learning algorithms make. To accomplish this, two common feature engineering functions are utilized in this workflow: a Z-score transform and the Yeo-Johnson transform. The Z-score transforms the data utilizing *eq. 2* where a given sample is represented by x, the mean of the feature is

represented by μ, and standard deviation of the feature is represented by σ. Essentially, this transformation scales the data with respect to the mean and standard deviation of a given feature's distribution and grades a sample based on the number of standard deviations it lies from the mean [6]. This determines what is called the Z score for a sample, which is then used as the sample's new position or value in the feature's distribution.

$$Z = \frac{x-\mu}{\sigma} \qquad \text{(eq. 2)}$$

The Yeo-Johnson transform has a similar goal, although it is more specifically focused on improving symmetry of a given data set [7]. The Yeo-Johnson transform is a modification of the Box-Cox transformation which allows for negative input values, something that the Box-Cox does not allow. Both Yeo-Johnson and Box-Cox use a parameter λ, which determines what direction the feature's distribution is currently skewed towards and how it will be reshaped accordingly. This parameter is typically estimated by calculating the maximum likelihood of each unique feature, independently, and then transforming that feature. The Yeo-Jonhson transform takes the following form (*eq. 3*):

$$\psi(\lambda, x) = \begin{cases} \{(x+1)^{\lambda} - 1\}/\lambda & (x \geq 0, \lambda \neq 0), \\ \log(x+1) & (x \geq 0, \lambda = 0), \\ -\{(-x+1)^{2-\lambda} - 1\}/(2-\lambda) & (x < 0, \lambda \neq 2), \\ -\log(-x+1) & (x < 0, \lambda = 2). \end{cases} \qquad \text{(eq. 3)}$$

Researchers have tested several data-preprocessing schemes for purposes of supervised learning on well logs [29].

2.4.2 Outlier Detection

Outlier samples will cause shifts in the distributions of unsupervised clusters, due to potentially noisy data. To remove this noisy data, an isolation forest algorithm is utilized to detect outliers.

The principal assumption of the isolation forest algorithm is that outlier samples are few and different from the rest of the sample set. Outlier samples mostly reside in lower-density regions of the sample space. Using this assumption, the isolation forest algorithm begins creating a forest-styled hierarchy where samples are grouped by common characteristics and are separated by more and more criteria as the forest gets deeper. The outlier samples, in theory, will remain at the beginning of the forest near the top because outliers are easy to isolate by portioning the sample space using feature thresholds [8]. Outliers remain at the top of the forest due to the fact that they will have less in common with other samples than actual signal data points. Once outlier samples, or contamination samples, are determined they are cast out of the data set. More information about outlier detection for well logs has been presented by Misra et al. [25].

2.5 Multi-Level Clustering

We apply clustering in two levels to generate the pseudo-rock types. Each level of clustering applies two distinct clustering methods. Comparison of the clusters generated using the two distinct methods based on the degree of overlap/similarity between the clusters enables the creation of final clusters for a given level. In our study, the first level of clustering generates two final clusters, while the second level generates five final clusters. Similar multi-level clustering has been applied for visualizing the carbon dioxide content in a injection reservoir for purposes of carbon geo-sequestration [26].

2.5.1 K-Means Clustering

For the purposes of predicting pseudo-rock types, this workflow utilizes two unsupervised clustering algorithms. The K-Means algorithm applies a process of partitioning a population of

N-dimensions in to *k* sets [9]. K-Means determines *k* initial cluster centers, where *k* is defined by the user, and then each cluster center is refined to be the mean of constituent samples within similar clusters [10]. Each sample is then assigned to its closest cluster center, and thus cluster label. This process is iteratively refined until there are no further changes in samples to clusters.

2.5.2 Spectral Clustering

The second unsupervised algorithm by which pseudo-rock type was predicted is spectral clustering. This method of clustering utilizes the eigenvectors of a matrix which has been derived from the distances between samples [11]. In the *d*-dimensional space, where *d* is the number of eigenvectors, the eigenvectors construct a geometric representation of the data which is then partitioned heuristically [12]. Using the eigenvectors in this manner allows to reduce the problem from graph partitioning to vector partitioning.

2.6 Cluster Validation

2.6.1 Silhouette Score

In order to evaluate the robustness of the generated clusters, methods of cluster validation were necessary to be incorporated into the workflow. This analysis utilizes two primary methods of cluster validation: silhouette score and comparison of different unsupervised algorithms. Several other evaluation metrics have been applied in the past on subsurface data for purposes of evaluating the robustness of the clustering results [26]. The silhouette score is a metric which is calculated after an unsupervised algorithm has been allowed to cluster a data set. This metric calculates the intra-cluster distance and the inter-cluster distance for every sample in the data set

[13]. These values are both averaged with respect to all samples within each cluster and the difference between the average values are used to generate a value ranging from -1.0 to 1.0. An illustration of a silhouette score plot is shown in *fig. 3*.

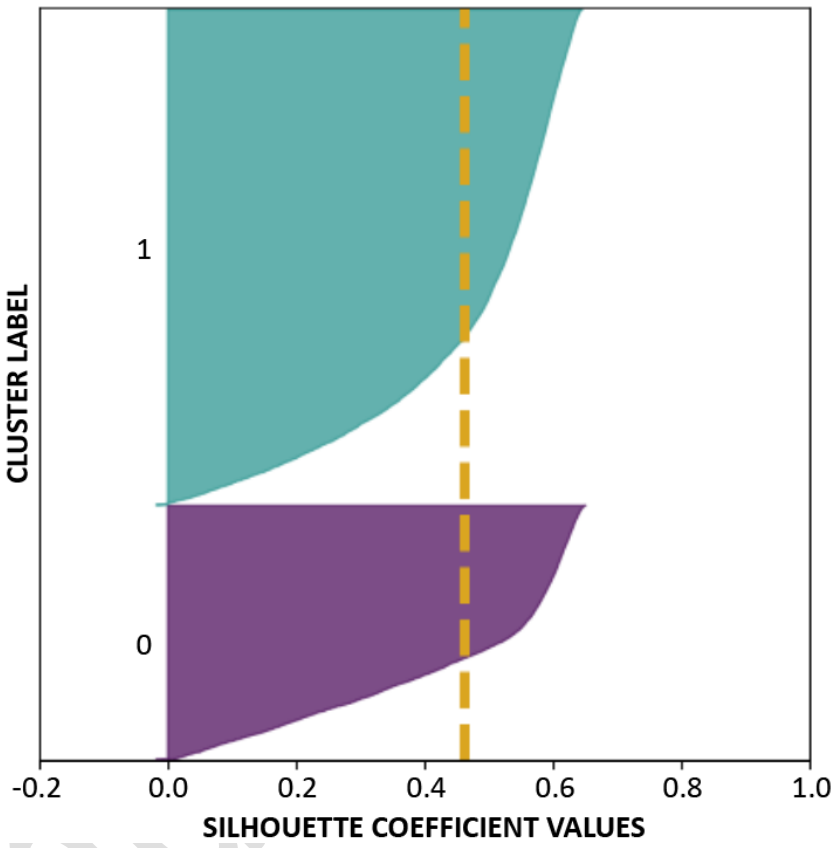

Fig. 3: Illustration of silhouette score plot, where the dashed yellow line represents the average silhouette score for the entire data set with the given clusters.

2.6.2 Comparison of Different Algorithms

Our second method of cluster evaluation is comparing different unsupervised algorithms. The two algorithms used to cluster the well log data are K-Means and spectral clustering. These two algorithms were allowed to cluster the samples simultaneously and separately. Once clustered,

the clusters can be visualized and compared side-by-side to compare cluster distributions in feature-space. It can be inferred that if two independent clustering algorithms based on distinct mathematical/optimization strategy generate a similar boundary between *n* number of clusters, it provides good indication that the clusters generated are robust signals and not randomly generated noise. This is illustrated in *fig. 4* where the results of K-means are compared to the results of spectral clustering for *n = 2* clusters in a reduced 2-dimensional feature space.

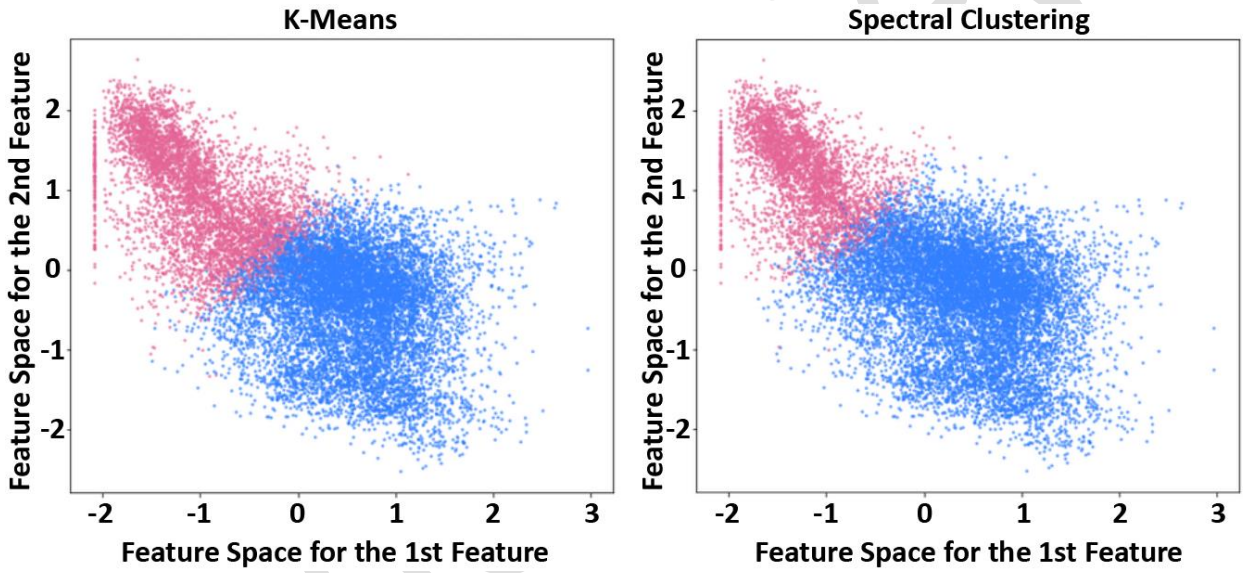

Fig. 4: Comparison plot of K-Means clustering and spectral clustering given n = 2 clusters, where the blue represents cluster 1 and pink represents cluster 2. The plot is presented on a reduced 2-dimensional feature space for visualization purposes.

2.7 Clustering Results

To begin applying the multi-level clustering, all 17 wells were treated as one continuous well; in other words, we mixed the data from all the wells. We had 1000 samples per well. Therefore, 5 log responses from approximately 17000 depth points (samples) were available for the clustering. For the first level of clustering, both the spectral clustering and the K-Means

clustering algorithms preferred to split the data set into two major rock types: A and B. Both of these two clusters contained a relatively large sample size, such that A has 10,000+ samples and B has 4,000+ samples. Out of 17000 samples, close to 1000 samples were removed because the two distinct clustering methods exhibited disagreement for these 1000 samples.

Clusters A and B were further subdivided into A0, A1, A2, B0 and B1 using the second level of clustering without re-scaling the features.. Based on silhouette scores and agreement between K-Means and spectral clustering algorithms, cluster A best divided into three sub-clusters. These clusters are named A0, A1, and A2. Similarly, cluster B divided into two sub-clusters: B0 and B1. The results of this multilayer clustering scheme are displayed for 5 HWPs and 5 LWPs in *fig. 5* below.

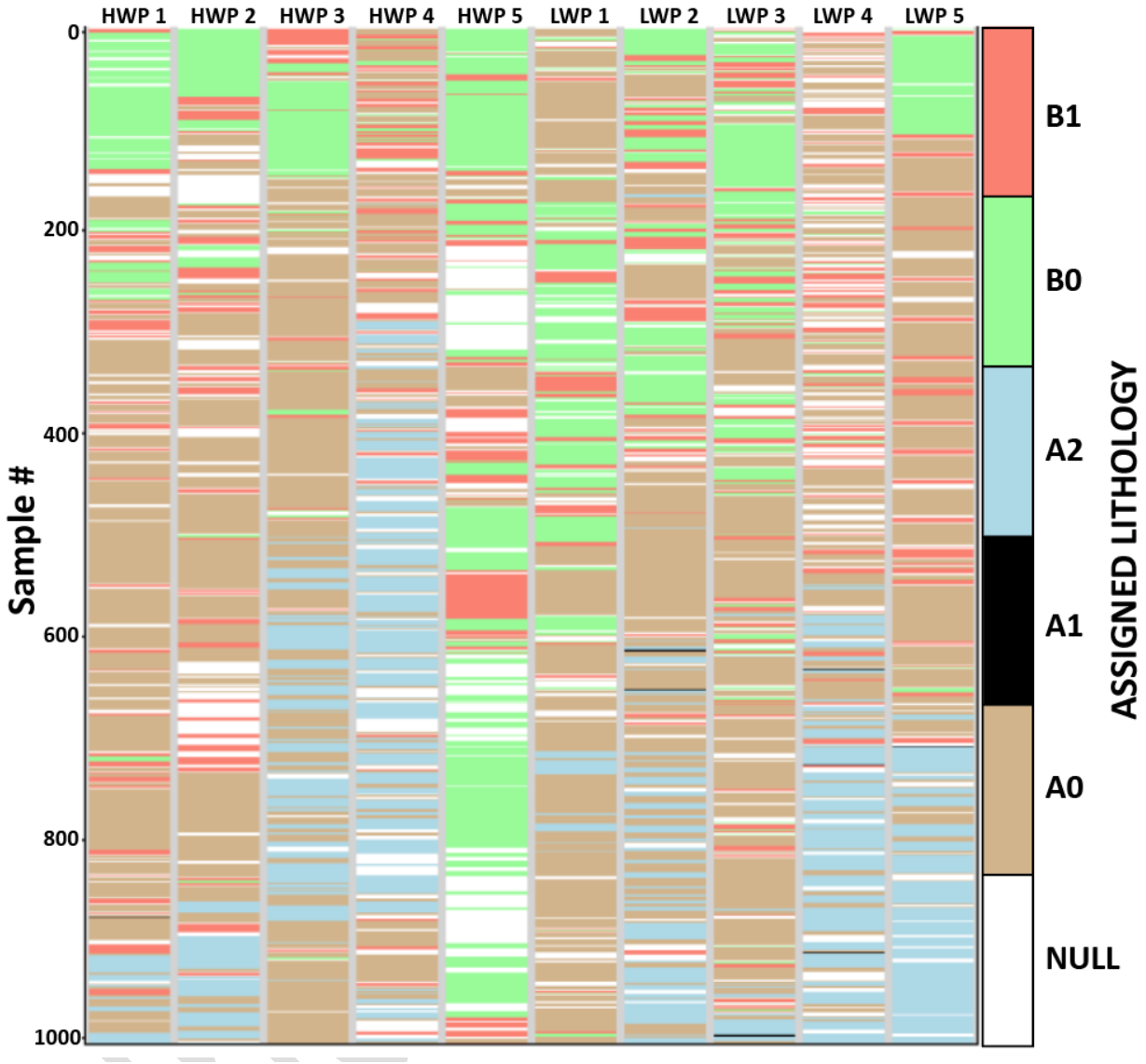

Fig. 5: Plot of the distribution of predicted lithologies from multilayer clustering process throughout the logged intervals for five high water producing wells (left half) and five low water producing wells (right half) randomly selected from the entire dataset.

2.8 Clustered Feature Extraction

Once all the samples were assigned a cluster, the feature extraction process began. The features for training supervised models needed to be defined on a well-by-well basis. Thus, the clustered

samples were redistributed back into their respective well. Features extracted took various forms. The first feature and the simplest was a binary feature which determined whether or not one of the 5 predicted clusters, or lithology types, was present in a given well. If this a cluster was classified as "present" in a well, the feature extraction process based on the given cluster was continued. If a cluster was not present in a given well, all subsequent features associated with that cluster are set to 0.

When a cluster is present in a well, a secondary check is applied for that cluster to determine whether or not there are at least 30 samples of this cluster in the given well. If this secondary check is passed, statistical summary parameters are extracted out from each well log with respect to the given cluster. These summary parameters consist of: mean, median, variance, kurtosis, root-mean square, skewness, and inter-quartile range. Each of these summary parameters were extracted to describe each well log within each cluster, granted there was sufficient sample count.

In addition to log-based features, depth-focused features were also extracted to determine if there was some correlation between relative depth of clustered lithology within a well and high water-cuts. In order to standardize depth measurements, depth was normalized with respect to each well's own KOP. Thus, the deepest depth is 300 ft from the KOP in the positive direction while the shallowest depth is 200 ft from the KOP in the negative direction. Depth-focused features were extracted on a cluster-by-cluster basis within a well, just as the log-based features were extracted. Similarly, statistical parameters were extracted based on relative depth measurements within a cluster. For example, the mean depth for cluster A1 would be calculated with respect to

a given well provided it had sufficient sample count of the A1 cluster. Each cluster's sample count was calculated, regardless if there was a minimum of 30 samples within a well, with respect to sliced depth intervals. The total depth interval was divided into five parts of equal thickness and each cluster's sample count was taken within that depth interval.

2.9 Feature Reduction

Once the feature extraction process was completed, there was a total of 255 features to use for supervised learning needed to relate the features to the target. Naturally, not all of these features will be useful for identifying the high water-cut wells. These features needed be reduced in order to optimize the predictive performance of the models which will be used to classify the wells in this data set. To accomplish this goal, three primary methods were utilized to reduce the feature-set: a One-Way ANOVA, Pearson's Correlation Coefficient, and Mutual Information score.

The first method, the One-Way ANOVA (analysis of variance) F-test is a standard statistical test which is capable of calculating correlation between continuous variables, such as most of the features in our data set, and discrete target labels, or our well classes. The F-test generates two linear regression models, where one is built with randomly selected constants assigned to a feature and another which one constant attached to a feature. If both of these models produce similar results, the null hypothesis is accepted and there is no statistical significance between the chosen feature and the target label [14]. This simple test generates two values which describe statistical significance: the p-value and the F-value. Both of these two values are highly correlated with one another and generally, a p-value $\leq .05$ and a F-value ≥ 1.0 are considered to

be statistically significant. However, in the context of this workflow only the p-value is utilized and the threshold which we reduce features by varies depending on an algorithm's prediction accuracy from the resultant feature set.

The second method utilized to reduce the feature set is mutual information (MI). MI is a metric which calculates the shared dependency between two variables, features and target. The MI between feature and target is calculated based on a joint probability density function between two variables and each variable's marginal probability density functions [15]. This is a metric which can be applied to both continuous and discrete variables, which is ideal as the data set contains both types of variables in the feature set. Put simply, a high MI value indicates high dependency between two variables and a low MI value indicates a more independent relationship between features. With this in mind, a threshold was applied to remove the features that have low MI score. The optimal MI threshold value varied depending on the supervised learning method.

The third method which is used for feature reduction is the Pearson's Correlation Coefficient (PCC). The intended result of filtering based on PCC is to reduce redundancy within the data set by removing features which are highly collinear to other features in the data set [16]. If two features are highly collinear, it can be assumed that they are sharing approximately the same information to the ML algorithms. The PCC generates a value ranging from -1.0 to 1.0, where 1.0 is perfectly collinear while -1.0 is perfectly collinear with a negative relationship. In the case of this workflow absolute values were utilized as we were only concerned with collinearity, as opposed to negative versus positive relationships.

2.10 Supervised Learning

2.10.1 K-Nearest Neighbors

This workflow utilized three supervised ML classifiers to preemptively identify the high water-cut wells: K-Nearest neighbors (KNN), support vector machine, and logistic regression. The KNN classifier utilizes the Euclidean distance between samples in feature-space to determine samples that belong to the same class [17]. This algorithm is dictated by the number of neighbors, *n*, which is specified to constitute a neighborhood within which the distances are calculated for purposes of assigning a class to a sample. Depending on which sample class has a "majority vote" within a neighborhood, this algorithm generates a model with a defined boundary used to assign that neighborhood to be descriptive of a sample class.

2.10.2 Support Vector Machine

The second supervised algorithm utilized in this workflow is the support vector machine (SVM). This algorithm is designed for two-group classification problems. With a SVM, input vectors are mapped non-linearly to a feature space with high dimensionality or many features [18]. Using this strategy, the SVM algorithm generates a decision boundary designed to maximally separate both sample classes. While this algorithm can be extended to create hyper-planes of high dimensionality, a linear decision boundary has proven to have the best results in this study because of the small dataset size.

2.10.3 Logistic Regression

The third supervised algorithm used to predict well classification in this workflow is the logistic regression. Modified from its predecessor the linear regression, the logistic regression is designed to produce binary outputs [19]. The logistic regression algorithm transforms continuous data to discrete classes using a sigmoid function bounded between values 0 and 1. The logistic regression model assigns a weight to each feature in the data set. Then the weights of these features are passed through the sigmoid function, where all samples will fall on a value between 1 and 0. A boundary is determined to separate each both classes which is then used to predict sample class based on the maximum log-likelihood distribution [20].

2.11 Training the Supervised Models

In order to better estimate the predictability of models generated, cross-validation technique was utilized. Cross-validation allowed us to determine the efficacy of produced models on various training and testing samples. Cross-validation takes the entire sample pool of a data set and splits it into separate, typically equal folds. The optimal number of number of folds determined for this approach was 5. With a 5-fold setup, the data from wells are split into 5 different folds. 4 of these 5 folds are used to train the supervised ML model, while the remaining fold is used to test the model's ability to predict well class. Owing the small dataset size, cross validation was performed multiple times to get a well-rounded statistical representation of how well these features can be utilized to identify the categories of water cuts, HWP vs. LWP. In this workflow, the cross-validation procedure is performed 100 times for each of the three supervised ML algorithms.

During each cross-validation iteration, the hyper-parameters of the given supervised algorithm are being optimized. These hyper-parameters are attributes of each algorithm which dictates the manner in which they learn. Optimizing the hyper-parameters was a key component to producing models which can most accurately predict the water cuts. The method used for hyper-parameter optimization in this workflow was the grid search method [21]. Using grid search, we provide the algorithms with a wide array of hyper-parameters to use in a series of trials. During these trials, every combination of hyper-parameters is tested to find which setup produces the best results. The best set of hyper-parameters was preserved after the trials and used to train and then predict well class from our features.

Each model generated from the supervised algorithms are evaluated using Matthew's Correlation Coefficient (MCC). The MCC score is a robust classification evaluation metric which quantifies a traditional confusion matrix into a numeric value ranging from -1.0 to 1.0, where 1.0 is a perfectly correct score and -1.0 is perfectly incorrect predictions [22]. The confusion matrix consists of four quadrants which describe predictions of a classifier: True positive (TP), true negative (TN), false positive (FP), and false negative (FN). The MCC score quantifies these quadrants using the following *eq. 4*:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad \text{(eq. 4)}$$

2.12 Feature Ranking Procedure

One of the primary goals of producing this workflow is to determine what underlying geological factors are contributing to the high water-cuts in the Delaware basin. In order to determine these

factors, it was necessary to investigate which features from the data set contributes the most information to our predictive models. To determine these most informative features, a process known as permutation testing is utilized [23]. This process takes one supervised model with known results and retrains and tests the model with an altered feature set. The altered feature set has every feature removed one at a time and the performance on each feature set is evaluated with the MCC metric. If a feature is important to the model's predictability, there will be a noticeable change in the model's MCC score when that feature has been removed through permutation testing. Several researchers have used extensive feature ranking schemes [27,28].

In order to get a wide understanding of what features are contributing to various models generated from these supervised algorithms, it was beneficial to perform permutation testing on models with varying MCC scores, or success rates. In order to achieve diversity in models, a resampling procedure from the models generated during cross-validation was utilized. A pool of 20 models were extracted from the 100 models generated during cross-validation. The median MCC score of these 20 models was calculated and compared to the median MCC score of the entire set of 100, which was 0.90. If the median MCC score of the set of 20 sampled models was greater than 0.85, the set of 20 was resampled until the median MCC score became less than 0.85. This limitation to the sampled pool was incorporated to ensure that there were less optimal models with more variety included in the sampled pool of 20 models. This process was performed for all three supervised algorithms to determine the most informative features when differentiating our two well classes.

The most informative features needed to be cumulatively ranked across many models and algorithms. In order to cumulatively rank them, they were first ranked model-by-model within each of the three respective algorithms. First, the features were ranked using a system used in our previous study [5] which incorporates ranked informativeness of a feature along with its frequency across multiple generated models. The following *equation 5* was used to rank features from the 20 sampled models:

$$R_f = \frac{\sum_i^n R_i}{f_i * 100} \quad \text{(eq. 5)}$$

where R_i represents the relative rank for a feature within one of the 20 models sampled for an algorithm, and f_i represents the frequency with which the feature displays significance through permutation importance, which is scaled by a factor of 100 to provide weight to features which occur more frequently across multiple models. The R_f represents the final rank within the set of 20 sampled models for one algorithm. Once a set of 8 to 10 most informative features was determined from a given algorithm, the same *eq. 5* was used to calculated a global rank across all three supervised algorithms.

3 RESULTS & DISCUSSION

3.1 Supervised Classifier Performance

The results produced by the 3 supervised algorithms are interestingly all comparable. Over 100 iterations of cross-validation, each of the 3 supervised algorithms produce an MCC score distribution with a median value of 0.90. Another point which should be mentioned is that the lowest MCC score for any models produced through the described methods is at least 0.70. This

provides confidence that the features produced through unsupervised lithology clustering convey important information for differentiating HWPs from LWPs. Granted that these features prove similar effectiveness for predicting HWPs and LWPs across different supervised algorithms, it can be assumed that the results produced from these three relatively simple ML algorithms can be improved with more complex learning systems with large dataset size is available. The performances of all three supervised algorithms are displayed in *table 1*, below. Both logistic regression and support vector machine outperform k-nearest neighbors by the narrowest of margins, when comparing their mean MCC scores.

Table 1. Results of supervised algorithm performances (Matthew's Correlation Coefficients [MCC]) for 100 cross-validation training and testing iterations.

Supervised Algorithm	Mean MCC Score	Median MCC Score
K-Nearest Neighbors	0.86	0.90
Logistic Regression	0.88	0.90
Support Vector Machine	0.88	0.90

3.2 Top 10 Ranked Features

The results from the permutation testing procedure are displayed in *Table 2*, where the top 10 most informative features across all algorithms are shown. Across all three algorithms the most informative feature was the kurtosis of the neutron porosity (NPHI) log within the A0 lithology. The second most informative feature to the supervised models is the presence of the A1 lithology. This feature is one of the binary checks that simply detect any presence at all of a sample labeled the given lithology. The third most informative feature to the supervised models is the kurtosis of the density porosity (DPHI) within the B0 lithology. Features ranked 4 through

10 are various features involving the B1 lithology. The difference in statistical significance between the B1 cluster's features and the top 3 ranked features is clear in the sense that every almost every model found the top 3 to be overwhelmingly significant. The exception to this would be the KNN algorithm, which did find the root-mean square of the gamma ray log to be ranked third.

Table 2. List of the top 10 most informative features for the preemptive detection of high water-cut wells, determined by the permutation testing procedure. Ranked from highest importance (#1) to lowest importance (#10).

Delaware Basin Cumulative Ranked Features for K-Nearest Neighbors, Logistic Regression, and Support Vector Machine	
Feature	**Rank**
A0_NPHI_Kurtosis	1
A1Present	2
B0_DPHI_Kurtosis	3
B1_GR_rms	4
B1icount3	5
B1_DPHI_rms	6
B1_GR_Variance & B1_ILD_mean	7
B1gr30	8
B1_ILD_median	9
B1_ILD_rms	10

3.3 Lithology Interpretation of Unsupervised Clustering

In order to better understand the physical meaning behind each of the 5 pseudo-rock type generated by unsupervised clustering, it was necessary to investigate the well log signature for each rock type. To best characterize the clusters, the centroids of each cluster were calculated. The centroid in this case was the arithmetic average of each well log within each cluster. It is important to note that the well log data in its final form is scaled and transformed by the means described in section *2.4.1*. This scaling at this stage of the workflow resulted in notable loss of

resolution. It was still possible to make some geologic interpretations of the pseudo-rock types based on their relative centroid in the three major axes of this workflow's utilized well log suite: porosity, resistivity, and gamma ray. The centroids along these major axes are illustrated in *Fig. 6*, where deep resistivity vs. average apparent porosity is plotted and the sizes of the points are scaled with respect to gamma ray signature. Other such 2D plots were tested but the one that generate a clear differentiation between the clusters is shown in *Fig. 6*.

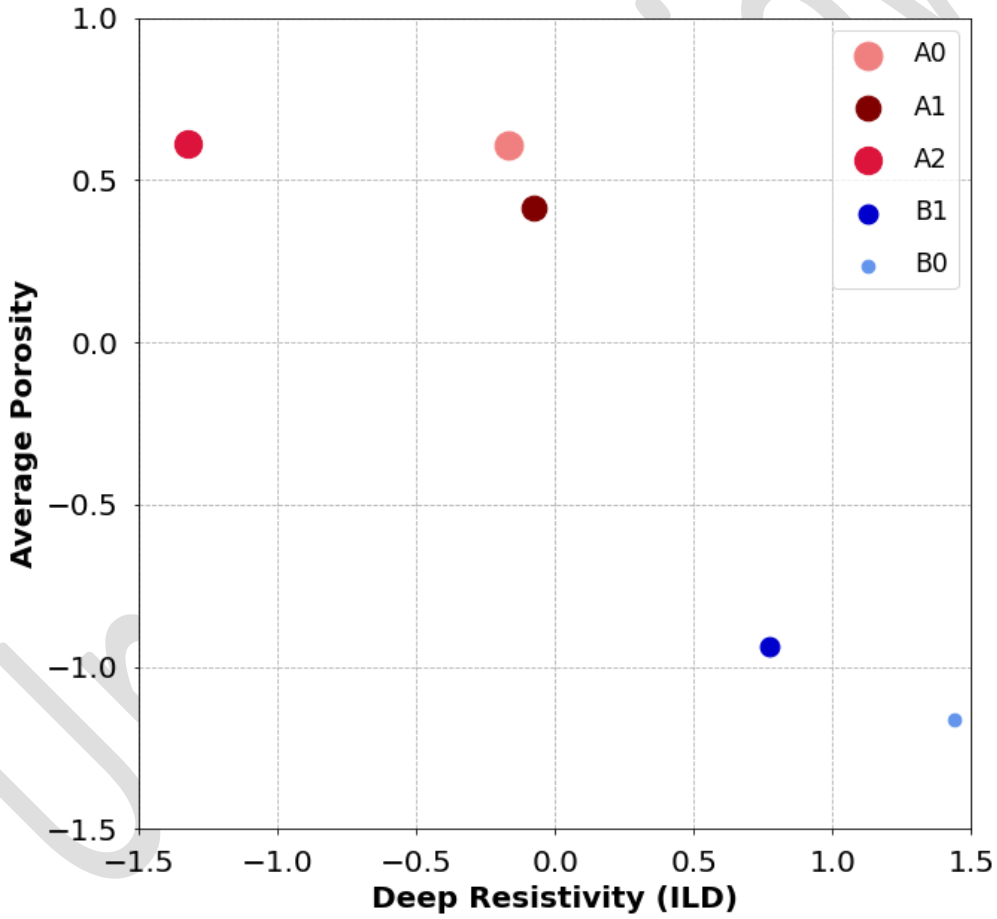

Fig. 6: Plot of the 5 predicted pseudo-rock types resulting from the multilayer clustering process. Size of the markers are scaled with respect to the gamma ray (smaller = lower GR signature). All features in this plot (Average porosity, deep resistivity, and gamma ray) are scaled with respect to the data set, where 0.0 represents the mean of each feature.

Starting with the original 2 principal clusters A and B, it is quite evident upon inspecting the results of *Fig. 6* that the multi-level clustering created clusters which have decent separation in all three of the major petrophysical properties in study. Cluster A can primarily be described as a lithology that has lower-than-average resistivity and high gamma ray signature with high apparent porosity. Neutron porosity values are large for shale rich formation because of large volume of clay-bound water and structural water in clays. Considering that our target formations are the bone springs sands and the Wolfcamp shale it is quite likely that cluster A can be considered a shaley rock which is relatively highly saturated with brine. Breaking this cluster down further, sub-cluster A2 appears to be much more heavily saturated with brine and conductive clays than any other pseudo-rock types. The low values of deep resistivity in this region are relatively indicative of brine saturation and clays as this will make the rock body more conductive than fresh water or hydrocarbon. The cluster A0 presents a very strong shale signal with its high GR signature, which is the highest out of all the sub-clusters of primary cluster A. Examining the resistivity signature from A0, it is likely that this lithology is the producing lithology of the Wolfcamp shale. The mean apparent porosity values for A0 are well above the mean of all the porosities in the sample pool. The A0 cluster is also the overwhelmingly largest out of all the predicted lithologies.

The A1 lithology is perhaps the most peculiar out of the 5 pseudo-rock types. The mean apparent porosity values of this lithology are not too different from lithology A0, but the resistivity and GR readings are significantly different. For every other predicted lithology, the deep resistivity is less resistive than the shallow resistivity, or nearly the same average value. This is not the case for the A1 lithology, in fact the opposite is the case. Within the A1 cluster, the mean shallow

resistivity value is nearly a standard deviation less than the average, while the deep resistivity is very close to the average expected value for deep resistivity of our samples. This separation between ILD and ILS could be indicative of a high permeability rock or mud-cake invasion into the formation. Though the invasion hypothesis is unlikely the case as using oil-based mud in the Delaware basin is by far the most common mud type. The presence of the A1 lithology being in the interval of interest is a strong predictor between HWPs and LWPs. Seven out of 9 low water producers have this lithology present in the interval of interest, so it is clear as to why this would be statistically significant. However, 2 out of the 8 HWPs in the data set also have this lithology present. Given that this lithology is mostly common in the LWPs, it provides evidence that this ILS and ILD separation could be indicative of a highly permeable oil producing rock. The gamma ray values for this lithology are just slightly higher than the average GR value expected amongst all the samples, so this is likely a dirty or shaley sandstone.

The primary cluster B seems to be characterized as low gamma ray, low apparent porosity, and high resistivity rocks. Differentiating between subclusters B0 and B1 is tricky, as there is significant separation in all three of the well log axes. The combination of very low gamma ray signature and high resistivity values characterizing the B0 lithology is typically indicative of a carbonate rock. The noticeable drop in porosity between B1 and B0 could also indicate that B0 is characterizing carbonates as certain carbonate rocks tend to deposit in thick beds with very little porosity. The gamma ray increase going from the centroid of B0 to B1 is likely indicative of a sandstone body. The increase in porosity as well could also be interpreted as a transition from carbonate to sandstone, however the porosity is still relatively low in comparison to cluster A. Generally speaking, the geologic interpretations discussed with respect to the centroids of these

pseudo-rock types are in agreement with generally agreed upon geology of the Upper Wolfcamp sections. The Upper Wolfcamp is generally described by geologists as a sequence of carbonate turbidites with varying total organic carbon (TOC) values [24].

3.4 Geologic Interpretations from Top Ranked Features

To get a better understanding of the physical interpretations of the top ranked features, the details of each needed to be further explored. In regards to the neutron porosity within the A0 lithology, the vast majority of the time the kurtosis for any given HWP is greater than any given LWP. This implies that the tails of the distributions of this well log extend out further for a given HWP, while the distributions for LWPs tend to be blockier. This means that generally the neutron porosity log reads more extreme values for HWPs, while the LWPs tend for NPHI values to stack up more centrally. Granted that the NPHI log detects fluid-filled porosity, this trend can be interpreted as these shales containing more deviations from the expected values of fluid-filled porosity. It is an intuitive interpretation that these outliers in saturated pore-space can result in high water-cuts. This could be refuted by the fact that these outlier NPHI readings could be indicative of hydrocarbon saturation, as opposed to water saturation, but with the contextual knowledge of already knowing that these wells are high water producers we can infer that these are likely water saturated samples. The difference between HWPs and LWPs is relatively large for this feature. The mean value of kurtosis for the NPHI log within the A0 lithology for HWPs is 0.17, while the mean value for LWPs is -0.31.

The second most important feature for differentiating high and low water producers is the presence of any samples of the A1 lithology within a given well. If this lithology is to be assumed to be a high permeability layer as interpreted in section *3*.3, these layers could function as conduits for water to flow through. However, these high permeability samples predominantly occur in the low water producing wells. Examining A1's distribution in *figure 6*, these samples generally occur deeper into the sampled well log region. Granted that the gamma ray signature in A1 is approximately half a standard deviation lower than the values of gamma ray in A0, this could indicate a dirty sandstone or perhaps even a thin sandstone interval. The latter of these two, the distribution of these samples would support due to their relative scarce distribution among the rest of the samples. Given that these layers most frequently occur in LWP wells, it is the most likely case that these are thin sandstone bodies which act primarily as conduits for hydrocarbons to be produced from.

The third most informative feature for the supervised models for the purpose of predicting well class is the kurtosis of the density porosity within the B0 lithology. Considering that the B0 lithology is most likely a carbonate signature, the interpretation of this feature is complicated. A clear trend displayed when examining *figure 5* is that this carbonate signature tends to be at the upper portion of the sampled log interval. The calculated KOP is at the 400-sample point (200ft) in *figure 5*, which provides insight that the vast majority of the B0 samples are stratigraphically above the producing interval in the horizontal wellbore. With this in mind, their influence to production is most likely associated with hydraulic fracture penetrating upwards through the rock bodies. The exception sample to this would be HWP 5 in *figure 5* where the horizontal wellbore also appears to predominantly be drilling through this presumed carbonate lithology. Similar to

the kurtosis of the NPHI within the A0 cluster, the vast majority of HWPs have higher kurtosis of the DPHI log within the B0 cluster. Thus, the distributions of DPHI are blockier within the LWP category than within the HWP wells. As the density porosity (DPHI) log is a product of a bulk density measurement, it is susceptible to influence by fluid inclusion as well. As either hydrocarbon or water saturation will lower the bulk density of a rock body, thus altering the density porosity to be higher, the outlier samples within a lithology can be interpreted as such. The mean value of kurtosis for the DPHI log in the B0 lithology is 0.57 for HWPs, while the mean value for the LWPs is -0.35.

The next 7 statistically significant features for the supervised models are all related to the B1 lithology. The B1 lithology is scattered about various relative depths within both HWPs and LWPs. Every well log within B1 seems to contain varying amounts of significant information for the supervised algorithms, with the exception of neutron porosity (NPHI). The B1 lithology likely represents a dirty sandstone or some type of clean rock with a relatively lower gamma ray signature than the shales (A0, A1, A2). With the very small sample size being worked with, the significance of features regarding this lithology type is unclear. Five out of the eight HWPs in the data set contain sufficient samples for statistical parameters to be extracted, the sample size greater than 30 samples requirement. Considering only the wells which meet this requirement, there are a few observations which can be made. All of the features represented of the B1 lithology in *Table 1* are on average lower for HWPs than for LWPs. The root mean square (RMS) of both the gamma ray and density porosity are both higher within the LWP wells. This is deceptive in regards to the DPHI, as the normalized porosity values within the B1 lithology are

all below the mean value and thus negative. The difference between mean DPHI values is still relatively significant, however the LWP wells have lower DPHI readings within this lithology.

4 CONCLUSIONS

In this analysis a data-driven workflow is prescribed to preemptively detect the high water-cut wells in the Delaware basin. This workflow utilizes unsupervised learning to predict pseudo-rock types from well log data. Once clustered, features were extracted from these pseudo-rock types and used to train supervised learning methods to differentiate high water cut wells for the low water cut wells. Using well log data from the Delaware basin, this workflow has produced promising prediction performances using multiple supervised methods. For 100 cross-validation training and testing iterations, a median Matthew's Correlation Coefficient of 0.90 has been generated for three supervised learning methods: K-Nearest neighbors, support vector machine, and logistic regression.

The pseudo-rock types identified by unsupervised learning broke down into 5 unique lithologies: A0, A1, A2, B0, and B1. The A group represented shalier rocks with higher gamma ray signatures and higher apparent porosities. The B group represented cleaner rocks, with lower porosity values. The A0 lithology likely represents the target shale formation for production as it has preferable porosity readings and high gamma ray signature. A1 was interpreted to be a high permeability zone characterized by a lower average shallow resistivity value than its average deep resistivity value. The A2 lithology was interpreted to be another shale formation with exceptionally low resistivity readings, which could represent a possible source of the water cuts.

Although no significant features were detected from the A2 lithology during permutation testing to find significant features to confirm this hypothesis. The B0 lithology was interpreted to be most likely a carbonate signature, due to low gamma ray and high resistivity readings. The B1 cluster most likely represents a sandstone or dirty sandstone formation as it is characterized by lower-than-average gamma ray, but higher than average resistivity.

The most informative features to the supervised models generated through this workflow were also examined. The most informative feature, by far, was the kurtosis of the neutron porosity log within the target shale A0 lithology. This is an intuitive result as the neutron porosity log specializes in fluid-saturated porosity calculations. Assuming that the kurtosis can be generalized as a number to represent the frequency of extreme values, it can be inferred that these extreme values of fluid-saturated porosity could be characteristic of high water-cut wells. The results of this analysis support this as high water-cut producing wells will on average have higher readings of kurtosis from the neutron porosity log. The second most informative feature to the supervised models is the presence of the A1 lithology. Generally, the wells which have this lithology within their sampled well log interval are going to be low water-cut wells. This high permeability lithology must act as a conduit for hydrocarbons to flow easier, assuming the results from this approach are to be trusted. The third most informative feature is the kurtosis of the density porosity log within the carbonate lithology, B0. The density porosity is a product of a transformation from the bulk density log. As bulk density can be affected by fluid inclusion, the only intuitive geological interpretation from this feature is that fluid inclusions are being detected with respect to HWP wells. HWP wells generally have higher kurtosis within the density porosity within the B0 lithology. Lastly the remaining features in the top 10 most informative

features are all associated with the B1 lithology. LWPs tend to have higher values of deep resistivity and gamma ray, but lower values of density porosity on average.

Acknowledgement

We want to thank Berg-Hughes Center for Petroleum and Sedimentary Systems and Crisman Institute for Petroleum Research at Texas A&M University for providing financial support for the project. Also, we thank Mark Nibbelink and his team in Enverus, who have helped us by providing access to data and consultations on technical aspects of the Enverus (formerly Drillinginfo) platform.

5 References

[1] Male, F. Using a segregated flow model to forecast production of oil, gas, and water in shale oil plays. Journal of Petroleum Science and Engineering, 180 (2019), 48-61.

[2] Duman*, R., 2019, October. Permian Produced Water: Impact of Rising Handling Costs and Larger Water Cuts on Wolfcamp Growth. In Unconventional Resources Technology Conference, Denver, Colorado, 22-24 July 2019 (pp. 4453-4460). Unconventional Resources Technology Conference (URTeC); Society of Exploration Geophysicists.

[3] Khan, N.A., Engle, M., Dungan, B., Holguin, F.O., Xu, P. and Carroll, K.C., 2016. Volatile-organic molecular characterization of shale-oil produced water from the Permian Basin. Chemosphere, 148, pp.126-136.

[4] Scanlon, B.R., Reedy, R.C., Male, F. and Walsh, M., 2017. Water issues related to transitioning from conventional to unconventional oil production in the Permian Basin. Environmental science & technology, 51(18), pp.10903-10912.

[5] Foster, J., Misra, S., Osogba, O., & Bhatia, M. (2021). Machine learning assisted detection of excess water-producing wells in unconventional shale plays. Journal of Natural Gas Science and Engineering, 104025.

[6] Jain, A., Nandakumar, K. and Ross, A., 2005. Score normalization in multimodal biometric systems. Pattern recognition, 38(12), pp.2270-2285.

[7] Yeo, I.K. and Johnson, R.A., 2000. A new family of power transformations to improve normality or symmetry. Biometrika, 87(4), pp.954-959.

[8] Liu, F.T., Ting, K.M. and Zhou, Z.H., 2008, December. Isolation forest. In 2008 eighth ieee international conference on data mining (pp. 413-422). IEEE.

[9] MacQueen, J., 1967, June. Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1, No. 14, pp. 281-297).

[10] Wagstaff, K., Cardie, C., Rogers, S. and Schroedl, S., 2001, June. Constrained k-means clustering with background knowledge. In Icml (Vol. 1, pp. 577-584).

[11] Ng, A., Jordan, M. and Weiss, Y., 2001. On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems, 14, pp.849-856.

[12] Alpert, C. J., & Yao, S. Z. (1995, January). Spectral partitioning: The more eigenvectors, the better. In Proceedings of the 32nd annual ACM/IEEE design automation conference (pp. 195-200).

[13] Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, pp.53-65.

[14] Osogba, O., Misra, S. and Xu, C., 2020. Machine learning workflow to predict multi-target subsurface signals for the exploration of hydrocarbon and water. Fuel, 278, p.118357.

[15] Estévez, P.A., Tesmer, M., Perez, C.A. and Zurada, J.M., 2009. Normalized mutual information feature selection. IEEE Transactions on neural networks, 20(2), pp.189-201.

[16] Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), pp.1157-1182.

[17] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), 21-27.

[18] Cortes, C. and Vapnik, V., 1995. Support-vector networks. Machine learning, 20(3), pp.273-297.

[19] Hosmer Jr, D.W., Lemeshow, S. and Sturdivant, R.X., 2013. Applied logistic regression (Vol. 398). John Wiley & Sons.

[20] Wu, Y., Misra, S., Sondergeld, C., Curtis, M. and Jernigen, J., 2019. Machine learning for locating organic matter and pores in scanning electron microscopy images of organic-rich shales. Fuel, 253, pp.662-676.

[21] Bergstra, J. and Bengio, Y., 2012. Random search for hyper-parameter optimization. Journal of machine learning research, 13(2).
[22] Chicco, D. and Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics, 21(1), pp.1-13.
[23] Pesarin, F. and Salmaso, L., 2010. The permutation testing approach: a review. Statistica, 70(4), pp.481-509.
[24] Kvale, E. P., & Rahman, M. (2016, August). Depositional facies and organic content of upper Wolfcamp Formation (Permian) Delaware Basin and implications for sequence stratigraphy and hydrocarbon source. In SPE/AAPG/SEG Unconventional Resources Technology Conference. OnePetro.
[25] Misra, S., Osogba, O., & Powers, M. (2019). Unsupervised outlier detection techniques for well logs and geophysical data. Machine Learning for Subsurface Characterization, 1.
[26] Gonzalez, K., & Misra, S. (2021). Visualization of the sequestered carbon-dioxide plume in the subsurface using unsupervised learning. https://www.essoar.org/doi/abs/10.1002/essoar.10507269.2
[27] Misra, S., & Wu, Y. (2019). Machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales with feature extraction and feature ranking. Machine Learning for Subsurface Characterization, 289.
[28] Ganguly, E., Misra, S., & Wu, Y. (2020, October). Generalizable Data-Driven Techniques for Microstructural Analysis of Shales. In SPE Annual Technical Conference and Exhibition. OnePetro.
[29] He, J., Li, H., & Misra, S. (2019). Data-driven in-situ sonic-log synthesis in shale reservoirs for geomechanical characterization. SPE Reservoir Evaluation & Engineering, 22(04), 1225-1239.