energies

MDPI

Article

Machine Learning-Based Probabilistic Lithofacies Prediction from Conventional Well Logs: A Case from the Umiat Oil Field of Alaska

Nilesh Dixit [1,*], Paul McColgan [2] and Kimberly Kusler [1]

[1] Physics, Geology and Engineering Technology Department, Northern Kentucky University, Nunn Drive, Highland Heights, KY 41099, USA; kuslerk1@nku.edu

[2] McColgan Seismic Interpretation Services, 7355 Huckleberry Lane, Montgomery, OH 45242, USA; mccolgan.avo@gmail.com

* Correspondence: dixitn1@nku.edu

Received: 29 August 2020; Accepted: 14 September 2020; Published: 17 September 2020



Abstract: A good understanding of different rock types and their distribution is critical to locate oil and gas accumulations in the subsurface. Traditionally, rock core samples are used to directly determine the exact rock facies and what geological environments might be present. Core samples are often expensive to recover and, therefore, not always available for each well. Wireline logs provide a cheaper alternative to core samples, but they do not distinguish between various rock facies alone. This problem can be overcome by integrating limited core data with largely available wireline log data with machine learning. Here, we presented an application of machine learning in rock facies predictions based on limited core data from the Umiat Oil Field of Alaska. First, we identified five sandstone reservoir facies within the Lower Grandstand Member using core samples and mineralogical data available for the Umiat 18 well. Next, we applied machine learning algorithms (ascendant hierarchical clustering, self-organizing maps, artificial neural network, and multi-resolution graph-based clustering) to available wireline log data to build our models trained with core-driven information. We found that self-organizing maps provided the best result among other techniques for facies predictions. We used the best self-organizing maps scheme for predicting similar reservoir facies in nearby uncored wells—Umiat 23H and SeaBee-1. We validated our facies prediction results for these wells with observed seismic data.

Keywords: machine learning; lithofacies; umiat; well logs; Alaska

1. Introduction

The success of any conventional hydrocarbon exploration program primarily depends upon identifying and mapping porous and permeable sandstone reservoirs where commercial volumes of hydrocarbons are stored. In these reservoir rocks, the spatial heterogeneity of porosity and permeability has been known to be affected by the geological character of the sediments, including their depositional environment, stratigraphic position, and mineral composition [1–3]. Detailed knowledge of sedimentary characteristics of high-quality sandstone reservoirs is, therefore, crucial in the early stages of hydrocarbon exploration.

Core samples are widely used in the petroleum industry for comprehensive analysis of lithologic and petrophysical properties of sandstone reservoirs that provide valuable insight into past depositional environments. Sandstones with consistent lithologic characteristics, such as grain size, appearance, and composition, are usually referred to as lithofacies [4]. Although core analysis provides rock information with high precision and spatial resolution, continuous coring is both an expensive and time-consuming process. Largely available conventional geophysical logs, including gamma-ray, density, neutron,

and resistivity, can be a cheaper and reliable alternative to predicting rock petrophysical parameters. However, these well logs do not provide direct observations of sand lithofacies and, therefore, cannot be linked directly to the depositional environments. Over the past ten years, research efforts are focused on the application of machine learning techniques to integrate conventional well data with limited core data, enabling reliable lithologic interpretations for sandstone reservoirs [5–7]. The key advantage of this machine learning-based approach is that it can be quickly applied to any number of nearby wells with limited or no core data once the data model is trained.

In this study, we demonstrated how advanced machine learning algorithms, including ascendant hierarchical clustering (AHC), self-organizing maps (SOM), artificial neural network (ANN), and multi-resolution graph-based clustering (MRGC), could be applied to conventional well log measurements from wells with no or poor core data to predict sandstone lithofacies. Our approach involved generating machine learning-based lithofacies models that were initially trained using limited core data and a suite of conventional well logs from a reference well (labeled data). These data models were then individually applied to well log data (unlabeled data) from nearby wells that had no core data to predict similar lithofacies. These predicted reservoir facies were then validated using seismic data. To demonstrate our approach, we analyzed the sandstones from the Lower Grandstand Member in the Umiat Oil Field—an emerging oil and gas frontier region of Alaska. Based on the limited core data available, these rocks are considered to have the best reservoir characteristics and are thought to be deposited by progradational wave-dominated deltaic river systems [8]. Five sandstone lithofacies were selected from the available core descriptions (761 ft to 868 ft) of the Lower Grandstand Member in an exploratory well, Umiat 18 [9]. These lithofacies were used to train the unsupervised machine learning models through machine learning techniques. We applied machine learning models to unclassified wells—Umiat 21H and SeaBee-1—for facies predictions, having only conventional wireline logs available and no core data. The resultant lithofacies prediction in two wells were validated through comparison with seismic data. Our results show vertical variability in deltaic sandstone facies observed within the Lower Grandstand Member, and they appear laterally continuous throughout the central portion of the field area. Our facies validation efforts further suggest that SOM is the best among the examined machine learning techniques and can be used as a supplementary tool to traditional core-driven interpretations of sandstone lithofacies.

2. Umiat Oil Field

The Umiat Oil Field of Alaska is one of the largest, undeveloped light-oil fields in the United States, with an estimated 1 billion barrels of original oil in place within the Cretaceous Nanushuk Formation [10] (Figure 1). The Umiat field was first discovered in 1946 and remains largely undeveloped over the past 70 years due to a lack of infrastructure at Umiat. Only 14 wells have been drilled into sandstone reservoirs of the Nanushuk Formation, with no significant oil or gas production. Previous studies suggest that the gas-oil-water contact exists in the Nanushuk Formation at about 1500 ft, with all sandstone reservoirs located within the hydrocarbon zone above the free-water level [11].

The Cretaceous Nanushuk Formation primarily consists of mudstone and coarsening-upward sequences of sandstone deposits that were deposited in fluvio-deltaic, shoreface, and marine-shelf environments [12,13] (Figure 2). Limited core data combined with outcrop observations show that the Nanushuk Formation has been subdivided into five lithologic units—the uppermost shallow marine Ninuluk sandstone unit, the deltaic Killik Tongue of the Chandler Formation, the underlying shallow marine to deltaic Upper and Lower Grandstand sandstone units that are separated by marine mudstones, and shales of the Tuktu [8]. Deltaic sandstones of the Lower Grandstand Formation are the primary exploration target in the Umiat Oil Field. Our study focused on the Lower Grandstand Formation, which is further comprised of two upward coarsening deltaic sandstone bodies that are separated by silty mudstones associated with lower offshore/shelf flooding surfaces (Table 1).

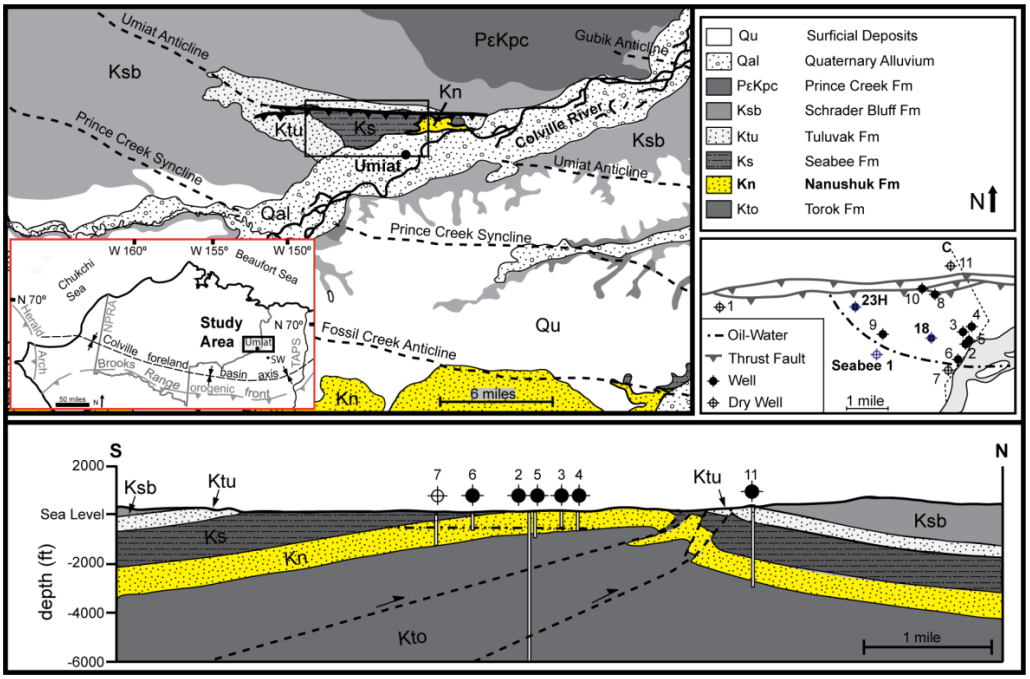


Figure 1. Reference geologic map, well locations, and a generalized cross-section of the Umiat Oil Field, Alaska. Modified from Shimer et al., 2014 [8].

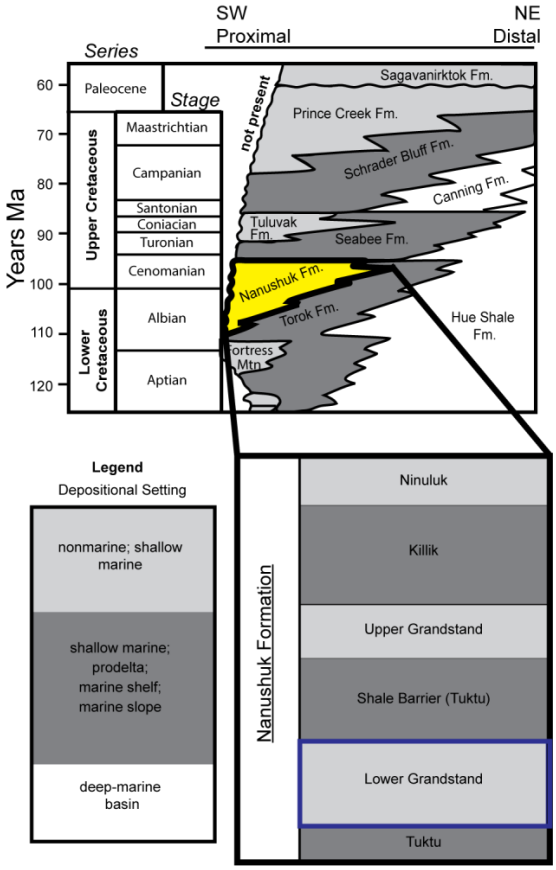


Figure 2. Stratigraphic column of the Nanushuk Formation, showing members and their depositional setting. Modified from Shimer et al., 2014 [8].

Table 1. The five lithofacies identified from cores of the Lower Grandstand Member in Umiat 18 (Modified from LePain et al., 2018 [9].

Facies	Description	Depositional Setting	Color
SI	Sandstone–horizontal, plane-parallel lamination	Distributary mouth bar and foreshore	1
Sx	Sandstone–Trough or planar cross-bedding	Foreshore and upper shoreface	2
Sm	Sandstone–massive	Delta-front and foreshore	3
Sr	Sandstone–ripple cross-lamination	Delta-front and lower shoreface	4
FI	Mudstone—carbonaceous	Distal self/prodelta	5

3. Data

In this study, we analyzed publicly available Umiat Oil Field data, including digital well logs, core descriptions, and seismic data. Digital well logs and well history files for three wells (Umiat 18, Umiat 23H, and SeaBee-1) were obtained from the Alaska Oil and Gas Conservation Commission (AOGCC) website (http://doa.alaska.gov/ogc/data.html). Digital logs include gamma-ray, neutron, density, sonic, and resistivity logs. Core data, including facies types [9], mineral composition, porosity, and permeability, from routine core analyses of the Lower Grandstand Formation, are also available publicly for Umiat 18 (depth range 710 ft to 1014 ft) on the website. Both prestack and poststack 2D and 3D seismic data within the Umiat Oil Field are available at the Alaska Geologic Materials Center (AGMC) (http://dggs.alaska.gov/gmc/seismic-well-data.php).

3.1. Lithofacies of the Lower Grandstand Member (LGST)

LePain et al. [9] examined cores (710 ft–1014 ft) from Umiat 18 and described five lithofacies within LGST that together comprise of prodelta and river-dominated delta-front facies (Table 1). These five reservoir lithofacies of LGST include SI- horizontal planer laminated sandstone, Sx- trough cross-bedded sandstone, Sm- massive sandstone, Sr- ripple cross-laminated sandstone, and FI- carbonaceous mudstone. These five lithofacies suggest a transition from an upper shoreface deltaic environment (SI, Sx, Sm, and Sr) to a distal shelf/prodelta environment. Porosity and permeability values follow subtle changes in lithofacies and are highest in high-energy facies of upward-coarsening deltaic-shoreface sandstones. Lithofacies data from Umiat 18 core were then used as a reference to train well log data from Umiat wells using machine learning methods.

3.2. Well Log Data

Geophysical well logs represent single-point measurements of rock physical properties with depth recorded in a well. A suite of conventional well logs from three wells (Umiat 18, Umiat 23H, and SeaBee-1) of the Umiat Oil Field were considered for machine learning-based lithofacies modeling and predictions. For each well, we utilized four scalar attributes from well log measurements, including:

- Gamma-ray (GR)—measures the total natural radioactivity emanating from a formation,
- Density (RHOB)—measures the rock bulk density based on the density of electrons in the formation,
- Neutron porosity (NPHI)—measures a formation's porosity by estimating neutron energy losses in porous rocks,
- Sonic (DT)—measures the travel time or velocity of an elastic wave through the formation.

Table 2 shows the depth range, minimum, maximum, and mean log values from all wells used in this study.

Table 2. The minimum, maximum, and mean values for input logs of the Umiat wells.

	Umiat 18			**Umiat 23H**			**SeaBee-1**		
Measured Depth (ft)	**10–2600**			**200–4100**			**100–15615**		
	GR	RHOB	NPHI	GR	RHOB	NPHI	GR	RHOB	NPHI
Unit	API	g/cm^3	V/V	API	g/cm^3	V/V	API	g/cm^3	V/V
Min	31.07	2.33	16.2	20.53	1.443	10	22.73	1.847	−0.09
Max	164.66	2.645	45.6	172.44	2.639	53	168.74	2.738	60.1
Mean	77.11	2.84	26.16	80.05	2.348	24	65.92	2.499	33.38

API—American Petroleum Institute.

4. Methods

4.1. Pre-Processing

Prediction of lithofacies using well log data relies on the assumption that each rock facies generates a unique set of log readings. In order to perform reliable facies prediction, all of the well logs need to be consistent and unbiased throughout the target lithological formation across all wells. Initial preprocessing of the original raw logs is, therefore, needed to remove any depth-related discrepancies, as well as any effect from borehole logging and the presence of hydrocarbons or other fluids.

We used the Paradigm^{TM} formation evaluation product—Geolog^{TM}—to perform initial log processing. All gamma-ray logs from Umiat wells were aligned in depth and plotted from 0 to 200 units. Calibrated GR values were then used for lithology predictions. The opposing response of neutron and density logs are commonly used to indicate the presence of hydrocarbons in the formation. Neutron porosity increases with increasing hydrocarbon content in the formation with a decrease in the formation's bulk density. If this hydrocarbon effect is not corrected, raw density and neutron porosity values can lead to wrong facies prediction in a well. We used neutron-density cross-plots to identify the hydrocarbon zones in LGST of Umiat 18. The identified hydrocarbon zones were analyzed in Geolog^{TM} to correct neutron and density logs for hydrocarbon effect corrections.

4.2. Machine Learning Algorithms

The workflow for the facies prediction includes four main steps: (1) selecting and pre-processing of training data, (2) building a machine learning model(s) to sort input data into clusters for facies prediction, (3) reviewing model(s) to identify an optimal number of clusters that fit the training data, and (4) applying model(s) to other wells in the area. For training purposes, we used three log attributes—GR, NPHI, and RHOB—as the model logs. We observed facies from core data as the reference log. We used DT as a reference log to create a synthetic seismic model for SeaBee-1. Initially, we applied machine learning workflows on the training data from LGST of Umiat 18 using Paradigm's Facimage^{TM} module. For each machine learning technique, we varied all combinations of the number of clusters and training parameters to generate different facies models for comparison and to identify the best matching scheme. We used R^2, the coefficient of determination, to analyze the goodness of fit between the observed and predicted facies. We employed the best machine-learning scheme to predict lithofacies from well logs in two un-cored wells—Umiat 23H and SeaBee-1—of the Umiat field. We examined the available seismic data to validate the predicted LGST lithofacies in test wells.

4.3. A. Ascendant Hierarchical Clustering (AHC)

AHC is an unsupervised machine learning algorithm used to find relatively homogeneous clusters of input data based on measured attributes [14,15]. This method employs a bottom-up way ("ascendant") to iteratively aggregate a pair of nearest clusters in order to have one final cluster containing all input variables. The method is outlined as follows:

Let's consider two data clusters, x and y, with an inter-cluster distance as d_{xy}. We assumed that cluster x contains N_i observations and cluster y N_y observations. If d_{xy} is the smallest distance that satisfies the conditions, (i) $x = y \Leftrightarrow d(x,y) = 0$, and (ii) $d(x,y) = d(y,x)$, clusters x and y are merged into a new cluster, p. Once a new cluster is created, the new minimum distance between p and any other cluster q is calculated using the Lance–Williams formula:

$$d_{pq} = \alpha_x d_{xp} + \alpha_y d_{yp} + \beta d_{xy} + \gamma \left| d_{xp} - d_{yp} \right|$$

where the values of α_x, α_y, β, and γ are calculated using Ward's minimum variance method [16]:

$$\alpha_x = (N_x + N_q)/(N_p + N_q),\ \alpha_y = (N_y + N_q)/(N_p + N_q),\ \beta = (-N_q)/(N_p + N_q),\ \text{and}\ \gamma = 0.$$

where N_x, N_q, and N_q are the number of elements in the clusters x, p, and q, respectively. For the minimum distance, d_{pqmin}, the closest two clusters are merged further into a new cluster. These steps are repeated until a single cluster with all observations remains.

For AHC, we gradually increased the number of classes into which the data was grouped from 10 to 50. At the beginning of the merging process, we considered each class as a separate cluster. Based on Ward's aggregation criteria, we iteratively merged the clusters until all the remaining clusters matched with the observed lithofacies classes.

4.4. B. Self-Organizing Maps (SOM)

Kohonen, in 1975, first introduced an unsupervised learning method that uses neural network-based algorithms for visualizing high-dimensional data in low dimensions. The SOM algorithm implemented in this study considers an adaptive learning process in which neurons are allowed to learn to represent different input data vectors. The neuron that best represents a selected input vector is considered as the winning neuron, and the neighboring neurons are gradually allowed to learn to represent similar inputs. Each neuron is then placed at the nodes of a one or two-dimensional lattice to transform multi-dimensional data into a low-dimensional discrete map.

This SOM-based data learning process can be expressed by the following equation [17]:

$$W_i \leftarrow W_i + \eta N(i,x)(x - W_i)$$

where W_i is the synaptic weight vector of the winning neuron ($i \in (0 \ldots \text{number of neurons})$), η is the learning rate that controls the size of the weight vector, x is the randomly identified input vector, and $N(i,x)$ is a neighborhood function that determines the rate of change of the neighborhood around the winning neuron. $N(i,x)$ is further defined as:

$$N(i,x) = \begin{cases} 1 \ for\ d(i,w) \leq \lambda \\ 0 \ for\ others \end{cases}$$

where $d(i,w)$ represents the Euclidean distance between the winning neuron and any ith neuron in the neighborhood, and λ is the iteration limit. It suggests that the neighborhood function returns to 1 for the winning neuron, which receives the most training, and is reduced to zero for neurons that are further away from the winning neuron, thereby receiving less training. These steps are repeated until the learning rates and neighborhood get smaller. When the neural network is fully trained, neurons that are similar are placed together, whereas neurons that are not similar are placed apart. The net result is a two-dimensional map of clusters, showing high-density nodes with similar inputs.

We trained SOM classifiers with a rectangular topography and with Euclidean distance as the distance function. The topology of the output layer was selected based on the training data and involved three schemes: 2×2, 5×5, and 7×7 neurons. In each case, the initial number of clusters were iteratively merged to five to match with the observed facies classes.

4.5. C. Artificial Neural Network (ANN)

A conventional back-propagation ANN algorithm is widely used to identify complex nonlinear relationships between input and output data, especially in pattern recognition applications [18–20]. ANN is a three-layer neural network structure with many interconnected nodes consisting of an input layer that collects the input data values, a hidden layer that performs the learning process via a system of weighted connections, and an output layer that shows the desired output, i.e., predicted patterns based on ANN's response to the input parameters.

The number of hidden layers and neurons in the hidden layers primarily influence the weighted connections and are, therefore, the most critical parameters in the ANN design. During the training phase of ANN, initial weights are usually randomly assigned between nodes in the first layer and are adjusted to minimize the differences between the actual and predicted outcomes. A weight decides how much the input value affects the layer output. An activation function is further applied to introduce nonlinearity by computing the weighted sum of input and biases at each neuron before passing the output to the next hidden layer. The resulting non-linear output is given by,

$$y = \alpha(w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n + b)$$

where x is the input, w is the corresponding weight to x, b is the bias value, and α is the activation function. When the backpropagation algorithm is used, the calculated error between the actual and desired output is propagated backward through hidden layers down to the input layer. Once the error for a hidden layer is known, the weights between the nodes are updated. This process is continued until the error is small enough across the hidden layers. When the model is trained to the desired level, the weight values are saved and stored before applying them to a new set of input data.

In this study, we implemented the ANN classifiers with one hidden layer and using a simple backpropagation algorithm. The neurons in the hidden layer were varied from 2 to 10 to examine the modes' prediction accuracy. We trained ANN models until we did not see any improvement in the model performance, compared to the observed data.

4.6. D. Multi-Resolution Graph-Based Clustering (MRGC)

Unlike conventional methods, MRGC is an unsupervised clustering technique that automatically calculates an optimal number of clusters independent of prior knowledge of input data. This method uses k-nearest neighbors and kernel index statistical parameters to identify the best number of dot clusters. Dots are representations of individual data points in a user-defined space to reveal data patterns in low-dimensions.

To develop a dot cluster, the proximity between each data point (x) to its nearest neighbor and other points in the space is identified using the nearest index (NI) [21,22] as expressed below:

$$\sigma_{k(y)} = exp(-\frac{p}{q})$$

where y is pth nearest neighbor of x among all data points, q is the total data points in the space, and $k = 1, 2, \ldots, K-1$ for x's K nearest neighbors. The sum of the rank of x relative to its pth nearest neighbor is expressed as:

$$S(x) = \sum_{k=1}^{q} \sigma(x)$$

The smallest sum of the limited ranks ($\sigma(x)$) is expressed as:

$$S_{min} = \min\{S(xi)\}$$

where i represents the number of measurement points from 1 up to K.

The largest sum of the limited ranks ($\sigma(x)$) is similarly expressed as:

$$S_{max} = max\{S(xi)\}$$

$$NI(x) = \frac{S(x) - S_{min}}{S_{max} - S_{min}}$$

Although *NI* provides a rapid estimate of nearest neighbor density at each data point in large data sets, it does not learn from the training data to do any generalization.

To overcome this issue, MRGC considers another function—kernel representation index (KRI)—which uses the weighted distance (M) and distance (D) between data points x and y to generate cluster indices. *KRI* at the point x is expressed by Rahimpour-Bonab and Aliakbardoust [21] as:

$$KRI\ (x) = NI(x) * M(x,y) * D(x,y)$$

where $M(x,y)$ is the number of neighbors, i.e., p, as y being pth neighbor of x, and $D(x,y)$ is the distance between x and y. If we consider m as any point from the nearest neighbors of x, and n is a point whose *NI* value is greater than m, then m merges all points around it, but it does not merge the point n. In this way, the kernel of a cluster is identified. When an appropriate kernel and its nearby neighbors are identified, cluster boundaries are set. Kernels are, therefore, localized in space and do not influence data points beyond the cluster boundaries. To obtain MRGC models, we considered the optimal model number to be between 3 and 5.

5. Results and Discussion

Our study aimed to examine the potential of different machine learning methods to predict lithofacies from well logs using limited core data from the Umiat Oil Field. Figure 3 shows the predicted lithofacies from the well log data of Umiat 18 using different machine learning methods, compared to the observed core-based lithofacies. Our results show that SOM provides facies predictions, closely mimicking the observed facies among all examined machine learning methods.

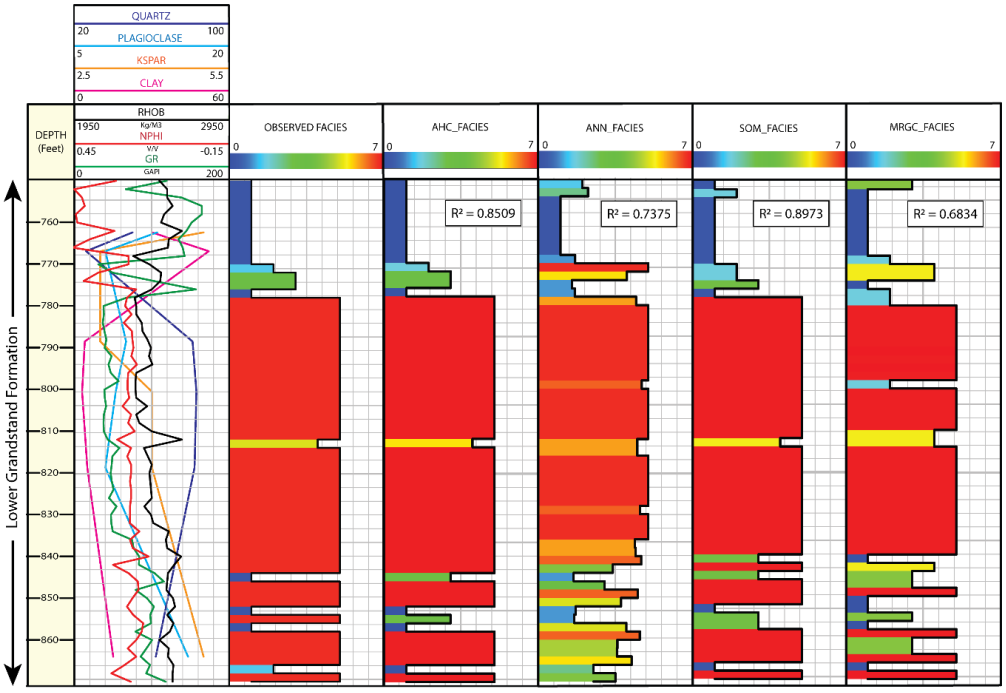


Figure 3. Comparison of observed vs. predicted lithofacies in Umiat 18.

5.1. Model Comparison

In this study, training sample data that we used to evaluate the performance of prediction models included 62 samples. Table 3 shows a range of parameters considered for a number of training trials for each machine learning technique. All models except MRGC were able to predict the lithofacies with a higher value of R^2 ($R^2 > 0.70$). Regarding the model performance, SOM was able to predict the facies with the highest accuracy compared to both AHC and ANN.

Table 3. R^2 values for each of the combinations of machine learning algorithm parameters used in this study. SOM with 7 neurons was observed to be the best among the examined methods.

Method	**AHC**			**ANN**			**SOM**			**MRGC**		
Factor	**Number of Classes**			**Neurons in Hidden Layer**			**Number of Neurons**			**Number of Optimal Models**		
-	10	25	40	2	5	10	2×2	5×5	7×7	3	4	5
R^2	0.52	0.75	0.85	0.65	0.73	0.74	0.75	0.79	0.90	0.33	0.64	0.68

Our model outputs further suggest the training with SOM tends to provide better results with increasing numbers of neurons. From a rectangular topology, the best SOM scheme with 7×7 neurons achieved a better clustering quality ($R^2 = 0.90$) (Figure 4, Table 3). It is, therefore, important to select the neurons in the output layer equal to the number of input classes from the training data. For AHC, the best result was obtained when the number of input classes was relatively higher and close to the number of samples in the training set ($R^2 = 0.85$). Figure 5 shows that the number of classes strongly influences the aggregation style and the resulting facies models. We examined ANN using a smaller training set (62 samples) and one hidden layer. The number of neurons in the hidden layer was varied from 2 to 10. When only two were considered in a hidden layer, the model produced the lowest goodness-of-fit to the observed data ($R^2 = 0.65$). Increasing numbers of neurons from 2 to 10 in a hidden layer did not significantly affect the model output ($R^2 = 0.76$). Table 3 shows the influence of the number of neurons per hidden layer on the model output compared to the observed facies. For a smaller dataset, we recommend using at least two neurons in a hidden layer to improve the accuracy of ANN. The best value of R^2 for MRGC was obtained using five optimal models with 17 clusters ($R^2 = 0.68$). The poor performance of the MRGC was due to the small number of training samples and the irregularity in the training data, especially the uneven distribution of sandstone facies in the cored section of Umiat 18.

5.2. Lithofacies Prediction and Validation in Test Wells

In order to predict the lithofacies in uncored wells—Umiat 23H and SeaBee-1—we applied the best SOM model to the well data from these wells. Figure 6 shows the predicted lithofacies for both wells. Our initial observations suggest that lithofacies identified from the core data of Umiat 18 also occur in Umiat 23H and SeaBee-1 but at different depths. In both wells, sandstone lithofacies show a strong vertical heterogeneity.

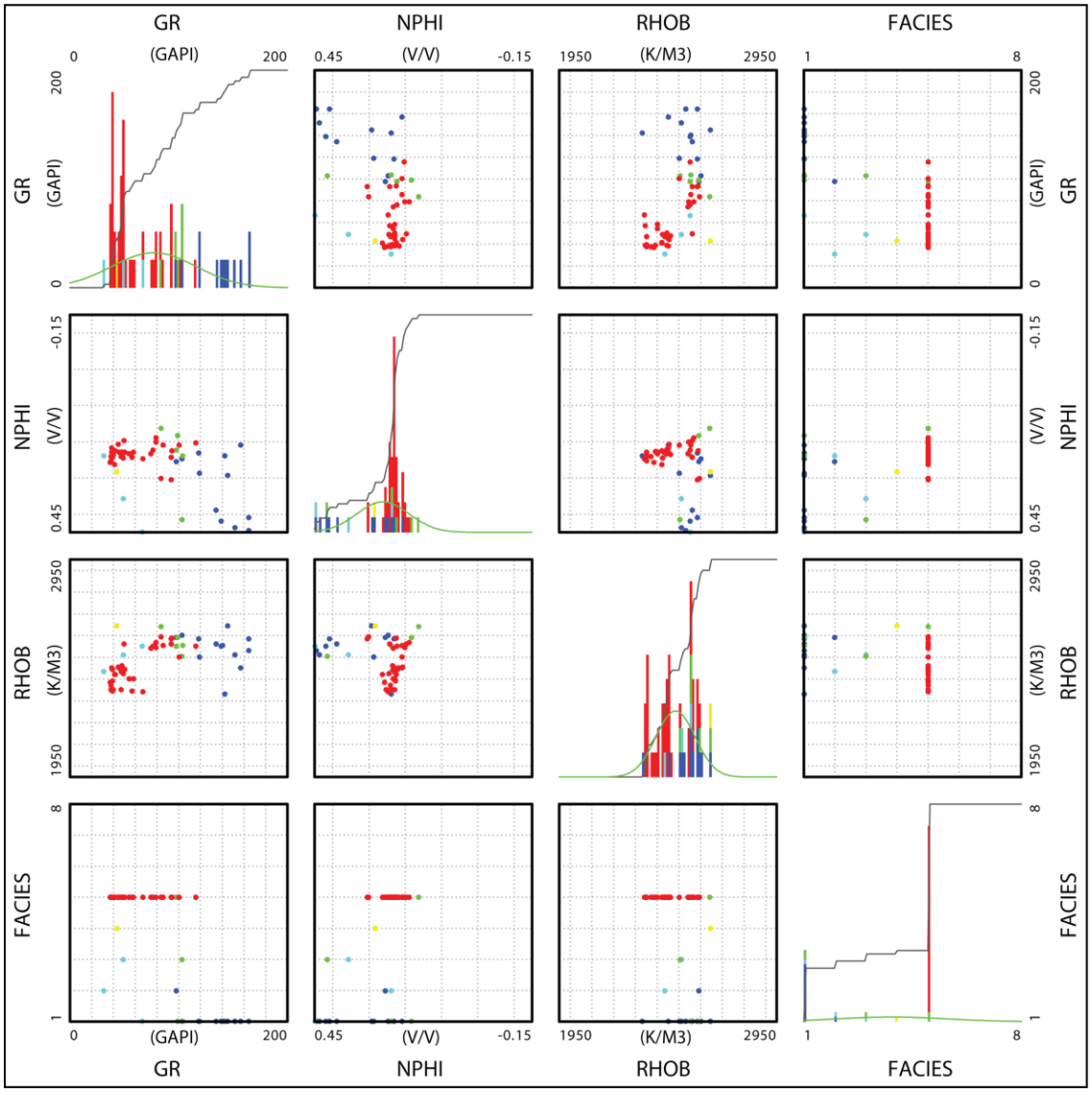


Figure 4. Distribution of self-organizing maps (SOM)-derived facies clusters and histograms, showing normalized training data. GR: Gamma-ray, NPHI: Neutron porosity, RHOB: Density.

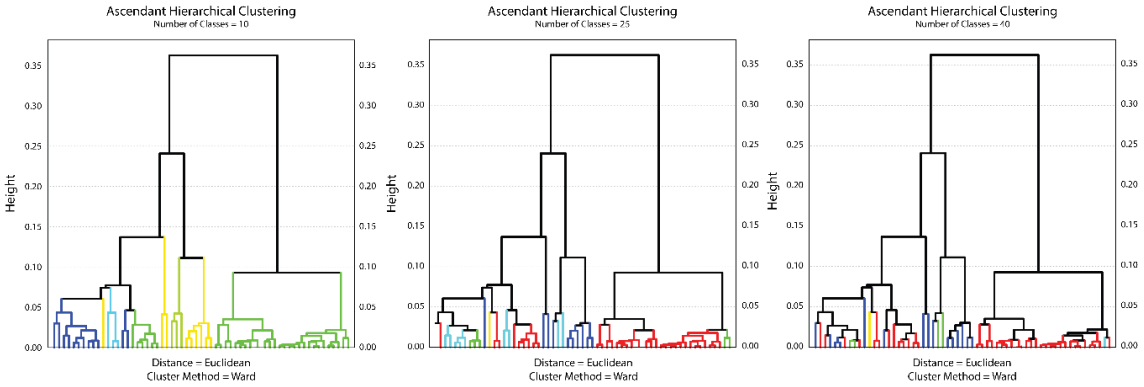


Figure 5. Figure 5 shows the dendrograms generated using a range of classes with Ward's aggregation criteria. Please refer to Table 1 for facies colors.

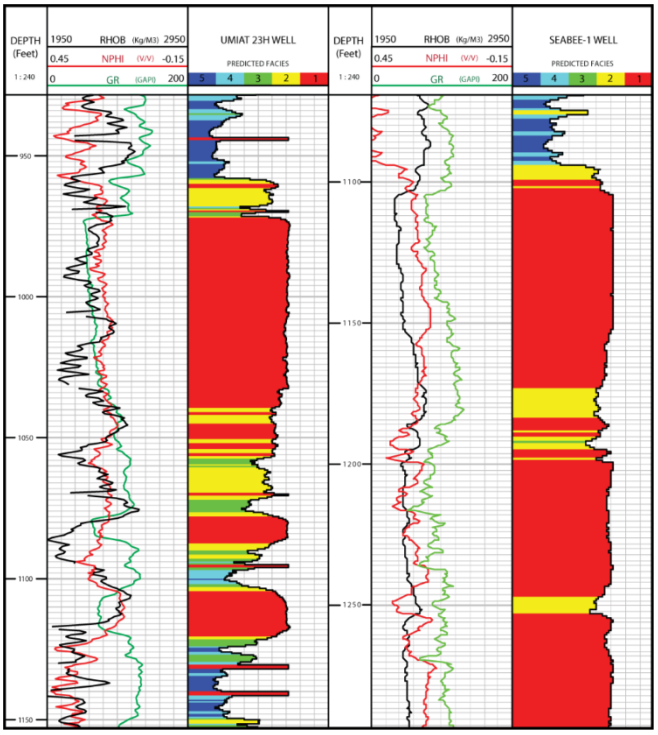


Figure 6. Umiat test wells (Umiat 23H and SeaBee-1), showing predicted lithofacies from the available well data using the SOM algorithm with 7×7 neurons.

To validate the presence of our predicted lithofacies, we examined the available seismic data across SeaBee-1. The predicted facies in Umiat 23H were not cross-validated due to the poor seismic data quality in the vicinity of the well. Figure 7 shows a synthetic seismogram constructed using sonic and density logs of SeaBee-1 in comparison to the predicted lithofacies for LGST in the well. Our predicted sandstone facies within LGST can be seen on the seismogram as negative (trough) amplitudes. To evaluate the hydrocarbon storage potential of these sandstone facies, we performed seismic amplitude variation with offset (AVO) analysis on the top sand in LGST using pre-stack seismic gathers. Figure 8 shows a strong AVO response along the horizon near the sand top. The computed background gradient for this particular set of gathers was −0.003, whereas the computed gradient for the displayed anomaly at the Lower Grandstand was −0.250, which was significantly higher than the background (Figure 9). The average stacked amplitude for the Lower Grandstand was roughly −300, whereas the stacked amplitude for the gather, exhibiting a strong negative gradient, was −625, which doubled the average value. The combination of a strong negative gradient in the pre-stack domain with doubling of amplitude in the stacked domain are strong indicators for a Class III AVO response at this particular Lower Grandstand location. These Class III AVO anomalies along the southern upthrown fault block of the Umiat anticline suggest a potential gas accumulation zone within the Lower Grandstand Member (Figures 2 and 10).

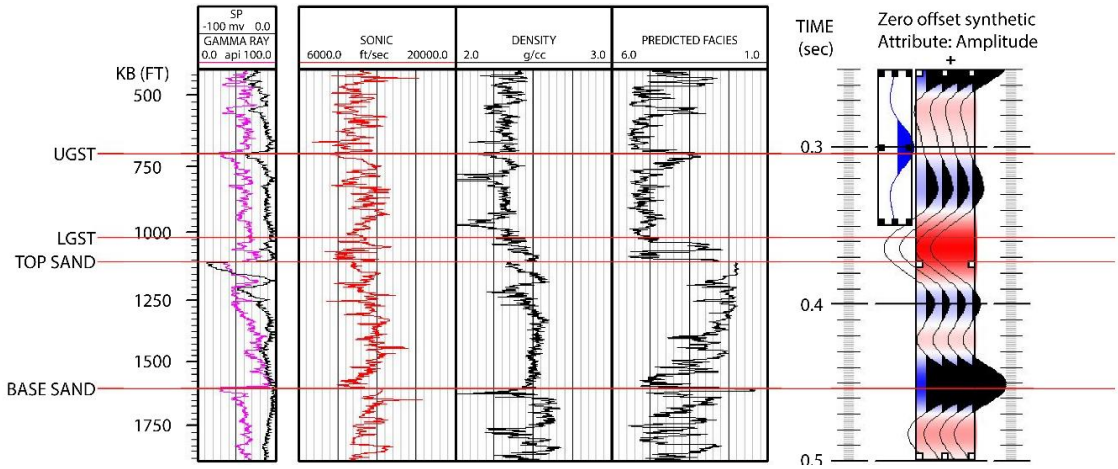


Figure 7. Correlation between synthetic seismic and predicted facies for the SeaBee-1 well.

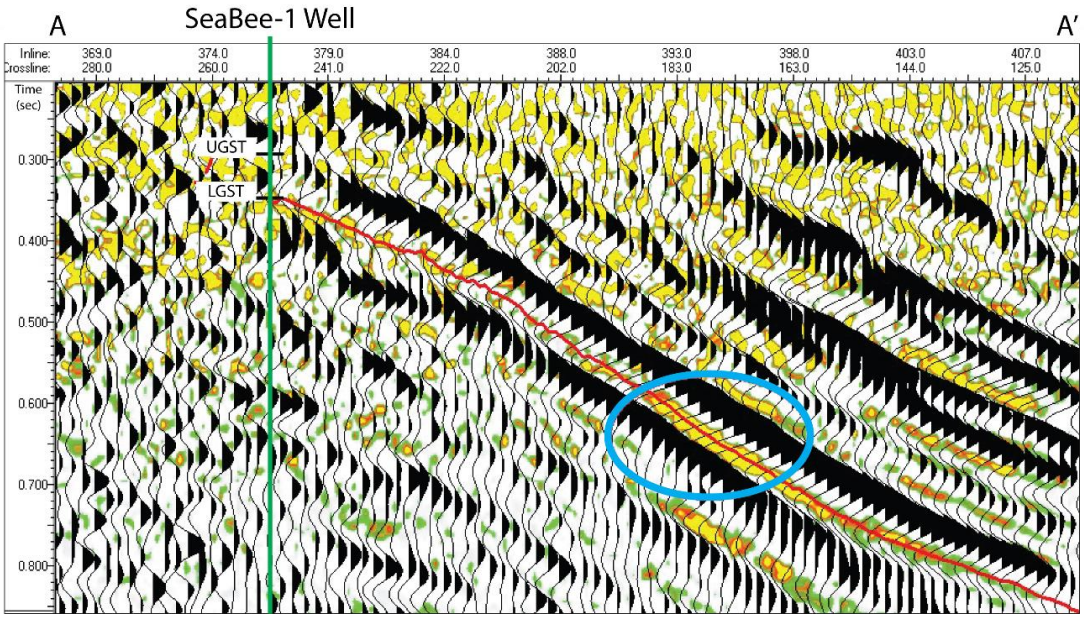


Figure 8. Seismic cross-section, representing the amplitude variation with offset (AVO) effect of the top sandstone reservoir facies along the horizon (marked in red) in the Lower Grandstand Formation.

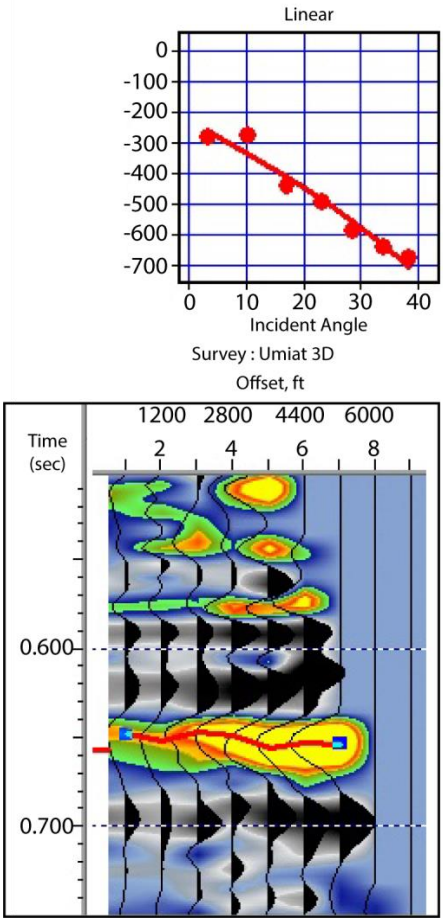


Figure 9. AVO intercept and slope, showing the presence of Class III gas sands within the Lower Grandstand Member at the Umiat Oil Field.

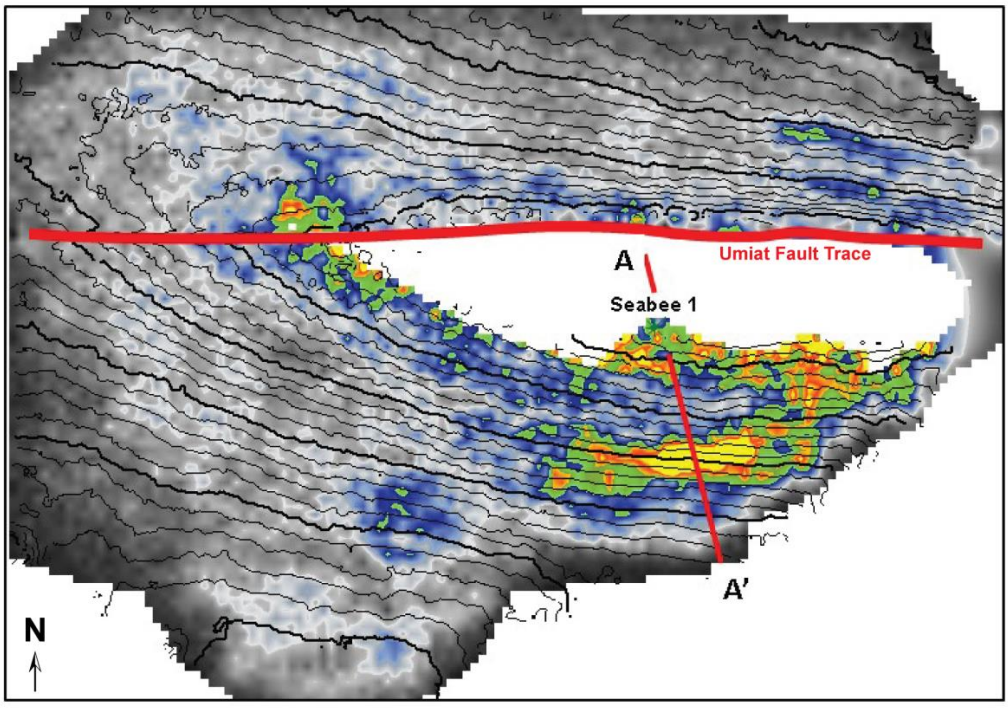


Figure 10. Time structure contours with AVO anomalies along the horizon, representing the top sand of LGST.

6. Conclusions

In our study, we demonstrate the application of four machine learning techniques—ascending hierarchical clustering, self-organizing maps, artificial neural network, and multi-resolution graph-based clustering, in order to predict the sandstone lithofacies from the well log data of Umiat wells. The following conclusions are drawn based on our work:

1. Machine learning techniques, such as AHC, SOM, ANN, and MGRC, can be used to integrate the facies descriptions from core data with the conventional geophysical logs for predicting lithofacies in uncored wells.
2. The predicted values of facies largely depend on the quality and size of input data, the machine learning method, and the training parameters used in machine learning algorithms.
3. We conclude that SOM is a better choice among all other methods for multi-dimensional well data with a small sample size. Despite lower model accuracy, the performances of ANN and MRGC can be significantly improved with more core data or training samples.
4. Application of machine learning techniques in the uncored wells can help in visualizing complex or multi-dimensional reservoir data in two-dimensions and can provide the assessment of reservoir quality at lower costs and saving time.

Author Contributions: N.D. and P.M. conceptualized and designed the research problem and methods. N.D. and K.K. performed the analyses of well data. P.M. processed the seismic data. N.D. wrote the paper. P.M. and K.K. reviewed the original draft of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We would like to thank NKU's Center for Integrative Natural Science and Mathematics for providing the research funding for this project. Special thanks to Emerson Paradigm™ for providing the Geology and StratEarth software programs to the Northern Kentucky University. We would also like to thank Grant Shimer from Southern Utah University for permission to modify and publish his figures.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Primmer, T.J.; Cade, C.A.; Evans, J.; Gluyas, J.G.; Hopkins, M.S.; Oxtoby, N.H.; Smalley, P.C.; Warren, E.A.; Worden, R.H. Global patterns in sandstone diagenesis: Their application to reservoir quality prediction for petroleum exploration. In *Reservoir Quality Prediction in Sandstones and Carbonates*; Kupecz, J.A., Gluyas, J., Bloch, S., Eds.; AAPG Memoir: Tulsa, OK, USA, 1997; Volume 69, pp. 61–77.
2. Johnston, D. Reservoir characterization improves stimulation, completion practices. *Oil Gas J.* **2004**, *102*, 60–63.
3. Taylor, T.; Giles, M.; Hathon, L.; Diggs, T.; Braunsdorf, N.; Birbiglia, G.; Kittridge, M.; Macaulay, C.; Espejo, I. Sandstone diagenesis and reservoir quality prediction: Models, Myths, and reality. *AAPG Bull.* **2010**, *94*, 1093–1132. [CrossRef]
4. Miall, A.D. Reconstructing the architecture and sequence stratigraphy of the preserved fluvial record as a tool for reservoir development: A reality check. *AAPG Bull.* **2006**, *90*, 989–1002. [CrossRef]
5. Dubois, M.K.; Bohling, G.C.; Chakrabarti, S. Comparison of four approaches to a rock facies classification problem. *Comput. Geosci.* **2007**, *33*, 599–617. [CrossRef]
6. Hall, B. Facies classification using machine learning. *Lead. Edge* **2016**, *35*, 906–909. [CrossRef]
7. Bestagini, P.; Lipari, V.; Tubaro, S. A machine learning approach to facies classification using well logs. In *SEG Technical Program Expanded Abstracts*; SEG: Houston, TX, USA, 2017; pp. 2137–2142. [CrossRef]
8. Shimer, G.T.; McCarthy, P.J.; Hanks, C.L. Sedimentology, stratigraphy, and reservoir properties of an unconventional, shallow, frozen reservoir in the Cretaceous Nanushuk Formation at Umiat Field, North Slope, Alaska. *AAPG Bull.* **2014**, *98*, 631–661. [CrossRef]

9. LePain, D.L.; Decker, P.L.; Helmold, K.P. *Brookian Core Workshop: Depositional Setting Potential Reservoir Facies, and Reservoir Quality in the Nanushuk Formation (Albian-Cenomanian), North Slope, Alaska*; Miscellaneous Publication 166; Alaska Division of Geological & Geophysical Surveys Miscellaneous Publication: Anchorage, AK, USA, 2018; p. 58. [CrossRef]
10. Hanks, C.L.; Shimer, G.; Kohshour, I.O.; Ahmadi, M.; McCarthy, P.J.; Dandekar, A.; Mongrain, J.; Wentz, R. Integrated reservoir characterization and simulation of a shallow, light-oil, low-temperature reservoir: Umiat field, National Petroleum Reserve, Alaska. *AAPG Bull.* **2014**, *98*, 563–585. [CrossRef]
11. Molenaar, C.M. Umiat field, an oil accumulation in a thrust-faulted anticline, North Slope of Alaska. In *Geologic Studies of the Cordilleran Thrust Belt: Rocky Mountain Association of Geologists*; Powers, R.B., Ed.; Rocky Mountain Association of Geologists: Denver, CO, USA, 1982; pp. 537–548.
12. Fox, J.E.; Lambert, P.W.; Pitman, J.K.; Wu, C.H. *A Study of Reservoir Characteristics of the Nanushuk and Colville Groups, Umiat Test Well 11, National Petroleum Reserve in Alaska*; U.S. Geological Survey Circular: Reston, VA, USA, 1979; Volume 820, p. 47.
13. Mull, C.G.; Houseknecht, D.W.; Bird, K.J. *Revised Cretaceous and Tertiary Stratigraphic Nomenclature in the Colville Basin, Northern Alaska*; U.S. Geological Survey: Fairbanks, AK, USA, 2003.
14. Lance, G.N.; Williams, W.T. A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems. *Comput. J.* **2003**, *9*, 373–380. [CrossRef]
15. Lukasova, A. Hierarchical Agglomerative Clustering Procedure. *Pattern Recognit.* **1979**, *11*, 365–381. [CrossRef]
16. Ward, J.H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [CrossRef]
17. Kohonen, T. The self-organizing map. *Proc. IEEE* **1990**, *78*, 1464–1480. [CrossRef]
18. Baldwin, J.L.; Bateman, R.M.; Wheatley, C.L. Application of a neural network to the problem of mineral identification from well logs. *Log Anal.* **1990**, *3*, 279–293.
19. Rogers, S.J.; Fang, J.H.; Karr, C.L.; Stanley, D.A. Determination of lithology from well logs using a neural network. *AAPG Bull.* **1992**, *76*, 731–739.
20. Bohling, G.C.; Dubois, M.K. An Integrated Application of Neural Network and Markov Chain Techniques to Prediction of Lithofacies from Well Logs (Kansas Geological Survey Open-File Report 2003-50). Available online: http://www.kgs.ku.edu/PRS/publication/2003/ofr2003-50.pdf (accessed on 17 September 2020).
21. Rahimpour-Bonab, H.; Aliakbardoust, E. Pore facies analysis: Incorporation of rock properties into pore geometry based classes in a Permo-Triassic carbonate reservoir in the Persian Gulf. *J. Geophys. Eng.* **2014**, *11*. [CrossRef]
22. Ye, S.J.; Rabiller, P. Multi-Resolution Graph-Based Clustering. U.S. Patent 6295504 B1, 25 September 2001.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).