Original Paper

Multi-attribute Selection for Salt Dome Detection Based on SVM and MLP Machine Learning Techniques

Nasrin Tavakolizadeh[1] **and Majid Bagheri** ◉[1,2]

Salt domes are one of the seismic patterns with exploration significance. This study focuses on supervised and unsupervised machine learning (ML) and feature selection methods using North Sea F3 block seismic data. Nine 3D dip-steered seismic attributes sensitive to chaotic features, including geometric, edge-detection, and two GLCM (gray level co-occurrence matrix) texture, were selected. A pickset consisting of 10,402 samples was gathered and normalized. Unsupervised self-organizing map (SOM), supervised support vector machine (SVM), and multi-layer perceptron (MLP) methods were applied through two different workflows. The dimensionality reduction techniques that were involved in the workflows included cross-plots, neighborhood components analysis (NCA), bagged decision tree, and Laplacian score (LS). SOM was trained by selected sections and the picked dataset. The latter elevated its performance and guided the neurons, differentiated salt, and noticeably reduced the misclassified samples. Learning curves were plotted to show the influences of different data populations on mean squared error (MSE). The results showed stability of SVM performance around 0.001 MSE for varying representation set size and fluctuation of MSE for the MLP method. For training the SVM/MLP, 35% of the pickset data was used. MLP and SVM with 99.79% and 99.9% accuracy, respectively, on the test dataset were applied to the sections. The supervised methods exploited the multi-attribute input and differentiated salt from the background. This study showed the importance of feature selection procedures and their resultant improvements in ML techniques. The workflows implemented here can be used in the automatic detection of seismic interpretation targets.

KEY WORDS: Salt dome, attribute, feature selection, self-organizing maps, support vector machine, multi-layer perceptron.

INTRODUCTION

Pinpointing seismic salt structures and their distribution is one of the essential steps in seismic data interpretation, and it has been investigated through various schemes. For instance, usage of texture attributes (Berthelot et al., 2013; Shafiq et al., 2017), resolution enhancement of seismic data

(Soleimani et al., 2018), or accelerating the process and automatizing the task by implementing unsupervised and supervised machine learning (ML) workflows (e.g., Buist et al., 2021; Di et al., 2018; Zhao et al., 2015). In this study, we focused on the latter methods. Unsupervised learning methods tend to find unknown patterns in a dataset. These workflows are advantageous because they work independent of a labeled dataset and unravel any cluster in data. Supervised learning using interpreter knowledge increases the accuracy and speed of facies interpretation by making it possible to gener-

[1]Institute of Geophysics, University of Tehran, Tehran, Iran.
[2]To whom correspondence should be addressed; e-mail: majid-bagheri@ut.ac.ir

alize an invariable outlook to an extensive database. The previous studies mentioned below have several areas of focus like applying new methods, improving the involved attributes, and some feature selection enhancement.

Bagheri and Riahi (2015) performed a multi-attribute facies classification study on an oilfield of Iran utilizing five supervised ML methods and four attributes selected by sequential forward feature selection (SFS), backward feature selection (SFB), and covariance matrix. They used well data as their training samples and the classified seismic lines to achieve seismic facies on the dataset. The SFS and SFB results were highly random due to the random selection of the starting feature. In the salt detection scheme, with the complexities of the classes, optimum feature selection can better be achieved by methods noticing all the features. Roden et al. (2015) concentrated on a methodology to extract existent patterns in the seismic data using a multi-attribute approach through principal components analysis (PCA) and self-organizing map (SOM). PCA performed dimensionality reduction for the SOM. They used the SOM for outlining stratigraphy and hydrocarbon indicators and distinguishing fractures and sweet spots. Zhao et al. (2015) compared selected supervised (including SVM (support vector machine)/MLP (multi-layer perceptron)) and unsupervised (including SOM) approaches for multi-attribute facies detection. The SVM as a deterministic method outperformed MLP. This study concluded that unsupervised learning with identifying unknown clusters could be a priori for supervised sample selection. Qi et al. (2016) stated that although salt is a chaotic structure, internal salt reflectors show some coherent noises or migration artifacts, which can impose a salt-and-pepper (mixed high- and low-value) look on the salt structure. Thus, they used Kuwahara median filters to precondition the attributes for the clustering step. By cross-correlation of their histogram, they selected relevant features. They performed generative topographic mapping (GTM) to assign probability density functions (PDFs) to each interpreted facies and then applied the trained device to an unlabeled cube. The Bhattacharyya distance between the resultant PDFs and the PDFs of the interpreted facies was stored in a probability volume, representing the user-defined facies. Although their results showed that Kuwahara filtering improved segmenting chaotic classes, smoothing results in attributes losing details especially on edges in seismic image. Di and AlRegib

(2017), employing twelve seismic attributes, tried to compare the performance of six supervised classification techniques in salt detection. They concluded that the ML methods could segment salt boundaries with reliable outcomes with a well-selected attribute set. Roden et al. (2017) studied identifying thin-bed or below-tuning layers utilizing instantaneous attributes and SOM; they declared that their SOM workflow recognizes the below-tuning clusters that are overlooked by other methods. Waldeland and Solberg (2017) tried to segment salt structures using convolutional neural net (CNN) and training a seismic image; they declared that by annotating an individual in-line as training data, it is possible to classify facies of a seismic cube. In fact, a CNN should be familiar with every geometry of possible pattern of facies in seismic images, and it needs a big and exhaustive interpreted training dataset and computational resources.

Di et al. (2018) implemented a multi-attribute unsupervised workflow using k-means clustering to detect 3D salt boundary. Manually picked samples were used to initiate the center of clusters, and the k-means clustering was trained to use the interpreter's knowledge and to start the clustering process more efficiently. The trained device was applied to the seismic volume and led to a probability volume showing the geometry and depth of each point on the salt dome. In this study, results were focused around the salt dome, and the most significant impedance variation was introduced as the salt boundary, and the attributes picked to focus on detecting it. Although salt boundary detection studies focus on finding sharp acoustic impedance contrasts, this is not always the case in real datasets. There are different complexities, for instance, the presence of multiple or no sharp boundaries. In addition, salt demonstrates various structures consisting of canopies, pillows, sheets, and walls (Fossen, 2016) and similar seismic patterns to mud volcanos or igneous intrusions. Using attributes that are sensitive to the chaotic texture of salt can be beneficial in detecting various structures. Chopra and Marfurt (2018) applied k-means clustering, PCA, SOM, and GTM to compare their results with their previous interactive interpretation on Barents Sea seismic volume. Although SOM and GTM provided promising results compared to the other two, their discrepancies included more smoothed and detailed channel facies distribution for SOM and GTM, respectively. Farrokhnia et al. (2018) started the salt classification process by implement-

ing a common reflection surface (CRS) on a migrated seismic 2D section to increase the number of stacked traces and, consequently, enhance the continuity of reflectors and better differentiation of them in the seismic image. Then, salt detector attributes were applied to the CRS image, and they used a mineral exploration approach to weigh them. A probability for each image pixel was calculated based on the linear relationship between the weighted attributes. These probabilities were used to plot the IGU (intrinsic geological unit) map, again a mineral exploration approach, to present the highest probability of salt as a binary output. The IGU and SVM results of conventionally stacked and CRS seismic images were compared, and CRS showed lower misclassified pixels. They stated that the quality of salt segmentation was directly related to the quality of the input seismic image. The results for top salt and marginal boundaries contained misclassified pixels and displayed dependency of the procedure on the implemented attributes and could be enhanced with the contribution of higher performance features.

Di et al. (2019) integrated 14 manually selected attributes by SVM and MLP in a fault detection workflow and incorporated edge-detection, geometric, and texture attributes. They also used a 1-layer CNN to extract 16 attributes automatically and used them as input to an MLP network to consider variations in the local pattern of reflections. They presented SVM/MLP efficiency in incorporating local seismic attribute sets in sample level and inhibiting effects of a less representative or wrong attribute by multi-attribute classification. In addition, the MLP outweighed the SVM in detecting faults due to its smoothing results, and CNN extracted attributes could reduce the misclassified samples. La Marca-Molina et al. (2019) employed five attributes to a horizon for delineating sand-rich lithofacies and architectures. Using SOM for their facies classification assists in the integration of the attributes. They showed that selecting capable attributes, with a powerful learning device such as SOM, along with using the horizons to follow the trace of the facies, improves locating the interpretation target. Kim et al. (2019) focused on non-redundant attribute selection methods for supervised learning based on a labeled dataset. Redundancy of attributes was examined based on correlation analysis between couple attributes as well as attributes and output classes. They concluded that redundant attributes do not always show a high correlation, and

features should be quantitatively ranked. Their results showed that attribute selection algorithms reduce computational costs and elevate classification performance by concentrating on a non-redundant efficient feature set. Qi et al. (2020) worked on substituting manual labeling the facies by Gaussian mixture models, determining a feasible number of features, and combining them for delineating chaotic structures. The suitable attributes for supervised chaotic facies delineation noted in this study included dip deviation, coherence, GLCM textures. Di and AlRegib (2020) applied MLP and CNN for salt boundary delineation and applied the devices to the synthetic 2D section of the SEAM dataset. MLP was applied by using nine attributes with different weights in the classification, and CNN was implemented by extracting patterns from seismic amplitude images. MLP could efficiently incorporate seismic attributes in salt detection but with higher false-positive rate (FP). CNN was not influenced by less illustrative seismic attributes and noise and artifacts in the dataset. Because CNN uses amplitude images to extract attributes, it needs to be trained on a big and varied interpreted and annotated seismic image dataset to extract the features. Furthermore, applying it in 3D data to consider a lateral structure variation for precise segmentation is computationally expensive. Buist et al. (2021) implemented a reservoir characterization process by a pre-stack time migrated 3D seismic data and correlation of the seismic attributes (structural and frequency-based) to petrophysical data by SOM, which detected facies equated to wireline log and core analysis data to locate porosity and permeability in the seismic sections.

This paper covers: (1) computing dip-steered 3D attributes with qualitative ability to differentiate salt structures; (2) gathering 3D-picked samples from the cube; (3) employing feature selection procedures; (4) application of an unsupervised workflow; and (5) application of a supervised workflow. Finally, we compare the performances, mention the possibilities for the future of the workflows, and draw conclusion from the results.

DATASET

In the southern North Sea, salt structures generate oil traps (Remmelts, 1996). The presence of Permian Zechstein irregular salt diapirs (with minimum depth at -1500 m and a maximum thickness of
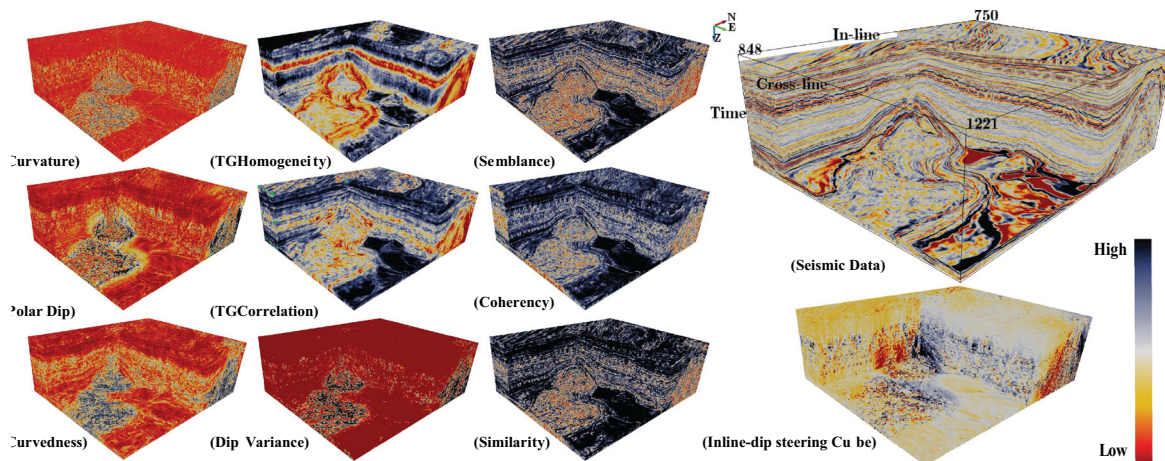
Figure 1. Seismic data, in-line dip-steering cube used to enhance attribute calculation by considering dip trajectories, and the calculated 3D seismic attributes in the study area.

700 m in the study area) dominates the eastern of the F3 block (Alaudah et al., 2019). The exposure region of salt diapirs in the Netherlands F3 seismic block, which was used in this study, is located between in-lines 100–750, cross-lines 697–1221, and time slices 848–1848. We selected some of the 2D lines of the 3D seismic cube to apply the learning techniques for illustrating the results and doing the process with lower computational cost.

Seismic Attributes

Salt structures distort seismic amplitude, and so attributes with the ability to detect this were employed, and nine features were selected after examining various ones based on their qualitative segmentation ability. The goal was to choose salt detector attributes that can be evaluated using dip-steered data cube (Fig. 1) to consider the amplitude distortions (Tingdahl et al., 2001, 2002), having a subset of comprehensive features with different nature (geometric, edge-detector, texture attributes), and calculating them in 3D. This 3D attribute calculation can allow learning algorithms to be deployed using any 2D line of the cube with one step of training an ML technique. The dip data are a dip trajectory that can be involved in attribute evaluations by giving a detailed view of the variations (Tingdahl, 2003). Positions of a selected phase are tracked and stored as a local dip and azimuth of

seismic events in the steering cube (Odoh et al., 2014; Tingdahl, 2003; Tingdahl & De Rooij, 2005) (Fig. 1). Four geometric attributes (dip, dip variance, curvature, and curvedness), three edge-detection attributes (coherence, semblance, similarity), and two texture attributes (homogeneity, correlation) were calculated and selected qualitatively for this study.

Curvature is sensitive to differences in fractured and deformed areas, and salt activity can be a source of prevalent deformation. Thus, two curvature attributes consisting of curvature (the most positive curvature) and curvedness were calculated (Barnes, 2016; Roberts, 2001). Curvature facilitates differentiating fractures in the seismic cube and maintaining their magnitude (Fig. 1). Curvedness values are positive and have geometrical importance (Fig. 1). It measures curvature independent from its shape, and coordinates and, because the measurements are positive values (Koenderink & Van Doorn, 1992), results accelerate a comprehensive evaluation of curved layers (Roberts, 2001).

Calculating true-time dips, polar dip, of time seismic data (Chopra & Marfurt, 2007; Marfurt, 2006) is a way to measure true geological dip (Fig. 1). Chaotic patterns like salt-related structures increase the possibility of dip variation. This can be evaluated by statistical dip measurements (Tingdahl, 2003), and volumetric seismic dip attributes have a functional role in the detection of small minor dip changes. Dip variance magnifies chaotic seismic
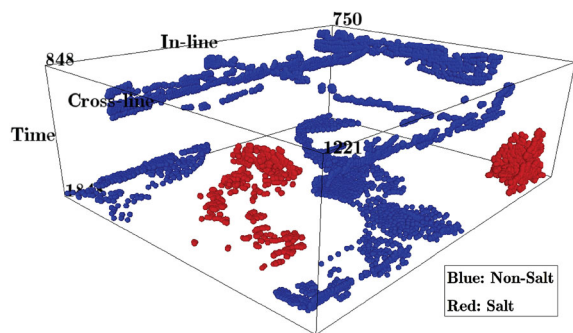
Figure 2. Distribution of 10,402 picked training samples. Red = salt. Blue = non-salt.

Figure 3. Histogram of the normalized pickset data sample values in the range 0–1.

features due to consideration of the statistical directionality of dip variation (Tingdahl, 2003) (Fig. 1). The input data used for calculating dip variance were the evaluated polar dip cube (Tingdahl & De Rooij, 2005).

Similarity attribute (Fig. 1) is measured by taking directionality into account through implementing steering cube data (Tingdahl, 2003). The distance between seismic trace segments at two different positions in hyperspace normalized to their length results in how similar those segments are. Salt chaotic patterns and seismic dip fluctuation can be distinguished by the coherence attribute. Lateral changes of these patterns in the cross-line and in-line directions are stored in a gridded coherence 3D volume (Bahorich & Farmer, 1995) (Fig. 1). The semblance (Fig. 1) is the ratio of stacked seismic trace energy to the average energy of each independent trace within the considered analysis window around a local planar event (Marfurt et al., 1998). Utilizing GLCM texture attributes ((TGH) and (TGC)) makes a quantitative interpretation of GLCM statistical figures achievable (Chopra & Alexeev, 2006; Eichkitz et al., 2015). Homogeneity is measured by weighted normalized GLCM elements, and it shows the similarity and differences of amplitude in neighboring samples (Chopra & Alexeev, 2006; Gao, 2007). Areas with internal irregularity (salt domes) are less homogeneous (Yenugu et al., 2010). Correlation is measured by the GLCM adjacent matrix's linear co-dependency, and the lowest correlations are accompanied by irregular beddings (Hall-Beyer, 2000).
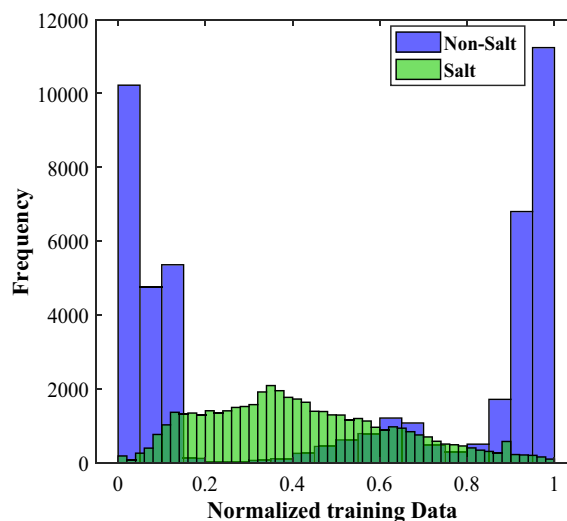
Pickset Data

In this study, a database of picked samples was gathered from the seismic cube. This dataset was selected from selected lines to avoid including training samples in the application procedure. The lines bearing the samples included in the pickset data were in-lines 394 and 697, cross-lines 1210 and 854, and time slices 1752 and 1810. The pickset data consist of 10,402 samples, including 5136 salt and 5266 non-salt data points (Fig. 2). The seismic cube used in this study contained more than 85 million samples, and the picked dataset was about 0.012% of this dataset. After selecting the samples, figures of the nine attributes were obtained at each point.

Sample data of every feature were normalized before entering the feature selection and training process. Normalization prevents overshadowing of features with low magnitudes by larger ones and contamination of results by outliers (Di et al., 2019). All the figures were distributed over the range from 0 to 1 (Fig. 3). The normalized dataset was used in feature selection processes and training the networks.

FEATURE SELECTION

ML results rely on training datasets, and a non-redundant illustrative feature set can enhance the results and reduce computational costs. In this sec-
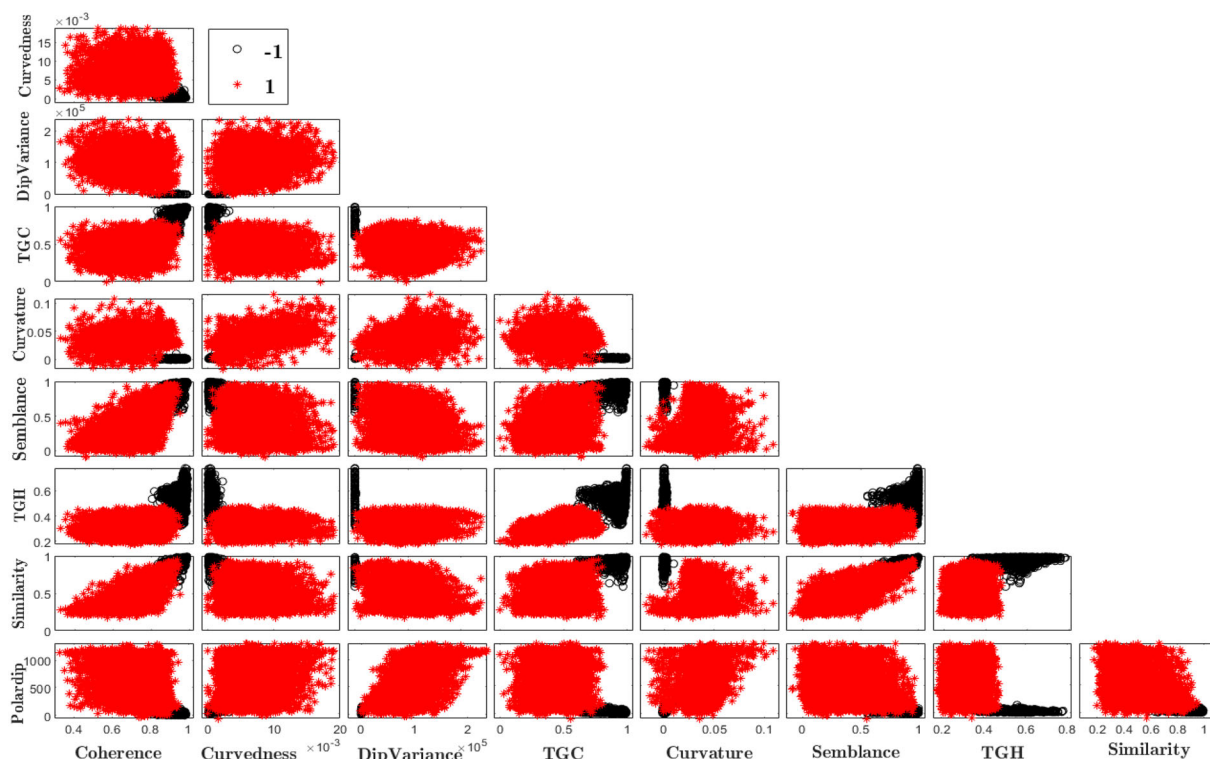
Figure 4. Cross-plots of nine qualitatively selected attributes. These plots show the capability of attributes in separating salt and non-salt classes. The greater the distance between classes, the higher is the performance of attributes in ML applications.

tion, four feature selection methods that were used in the supervised and unsupervised workflows are presented.

Cross-plots

Cross-plots, being a clustering technique, are created to assess redundancy of attributes. The salt and non-salt samples of each attribute were cross-plotted to measure the redundancy of each couple of attributes. Cross-plots with the most considerable distance between classes show which attributes are non-redundant (Qi et al., 2020). In contrast, the decrease in this distance and overlap of samples happen because of subsurface complex geology and noise in the dataset (Di et al., 2018) and these cause misclassification errors when applying classifiers (Di et al., 2019) and redundant features. In the cross-plots (Fig. 4), curvature–similarity, TGC–TGH, dip variance–TGC, and TGH–coherence showed the lowest overlap among the classes.

Neighborhood Component Analysis (NCA)

NCA is a nearest neighbor-based weighting algorithm. It learns the weights of a feature set by eliminating the irrelevant ones to reduce the leave-one-out cross-validation error with tuning the regularizing parameters using gradient ascent (Yang et al., 2012). It takes the class labels vector and selects features with better performance, classifying them starting from a reference random point determined by a probability distribution in a dataset and goes on to its nearest neighbor. Higher weights represent more important features in the classification task. Because NCA results have a random nature, we repeated the weighting method 10 times and the average of the final results is presented here (Table 1). According to this method, the weights showed that curvature and TGC were the top two attributes for identifying salt.

Table 1. Results of the NCA and bagged decision tree feature selection methods

Attributes	NCA	OBPI*
Polar dip	1.53	0.41
Curvedness	1.94	0.58
TGC	**2.87**	0.42
TGH	2.31	0.28
Semblance	1.49	0.24
DipVariance	2.12	**0.95**
Similarity	1.73	0.32
Coherence	1.34	0.33
Curvature	**3.04**	**0.59**

The highest values are in bold
*Out of bag predictor importance

Bagged Decision Tree

Decision trees with a simple arrangement can grade the importance of features in a hierarchical structure (Rounds, 1980); they consist of a root that includes the whole population of features and leaves that contain predicted responses. Through the tree and passing from each leaf, the significance of features in the classification is checked (Breiman et al., 1984).

Here, the bagged decision tree was used to improve the accuracy. The method results in a more accurate feature set by averaging over random bootstrap selections of the training data sets and by applying decision trees over them. An equal amount of data is used per tree in the bagged version (Breiman, 2001). Finally, the importance of the features was estimated by averaging over the ensemble of trees, and out of bag predictor importance of the features was evaluated (Table 1). These indices show that absence of a certain feature can cause more misclassified samples. The resultant tree (Fig. 5) shows that dip variance and curvature were the most influential attributes based on this method.

Laplacian Score (LS)

LS is an unsupervised feature selection technique introduced by He et al. (2005), and it is independent of a ML approach or a labeled dataset. This method relies on the fact that, in a dataset, close samples belong to similar classes. Where a dataset (X) consists of n samples and r features (f), the LS is assigned to the features in order to prioritize them based on their ability to preserve locality in the dataset. The method specifies the local
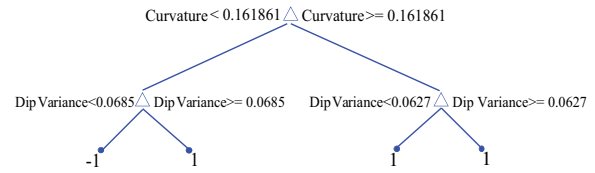

Figure 5. Final result of bagged decision tree applied on the sample set (1: salt, − 1: non-salt). Each branch classifies the data by contribution of the most dominant features; here, dip variance and curvature.

neighborhood for each data point in the dataset X ($= \mathbf{x}_i, \mathbf{x}_j, ..., \mathbf{x}_n$) by constructing a nearest neighbor graph G. This graph contains the same number of nodes (n) as the samples of the dataset. If a node in this graph ($\mathbf{x}_i$) belongs to the k nearest neighbors of another node ($\mathbf{x}_j$) (i.e., they are close to each other), they are connected, and the distance between them is calculated; and if nodes are outside the neighborhood, a weight $S_{ij} = 0$ is assigned to the pair. This weight allocation results in the neighbor graph's weight matrix, also called the similarity matrix (S). LS pursues the features to preserve the structure of this graph. Utilizing the degree matrix ($D = diag(S)$), each feature (f_r) is centered ($\sim f_r$) and using graph Laplacian matrix ($L = D − S$), the Laplacian score (L_r) of the rth feature is calculated as (He et al., 2005):

$$L_r = \frac{\sim f_r^T L \sim f_r}{\sim f_r^T D \sim f_r} \tag{1}$$

The scores for each of the attributes are presented in Table 3. The higher scores are the more critical in preserving the structure of the neighbor graph and in segmenting the classes. Unlike the NCA, the results in this method are stationary, and so we applied it to the pickset data used in supervised and unsupervised learning.

MACHINE LEARNING METHODS

In this study, one unsupervised and two supervised learning techniques were implemented in specific workflows. The learning devices and their workflows are discussed in this section.

Unsupervised Learning

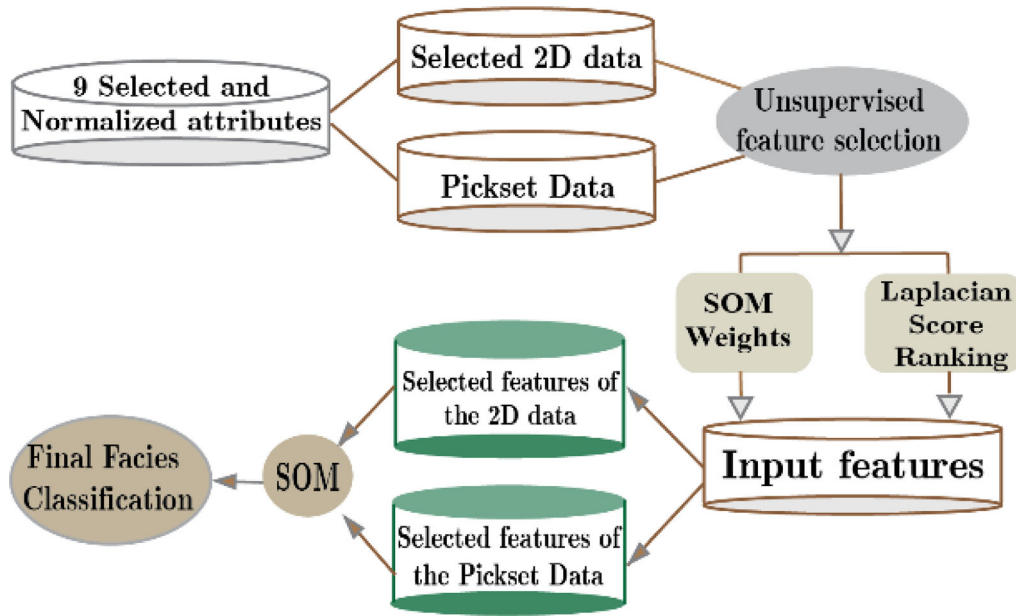Unsupervised ML techniques can be beneficial while investigating natural unknown patterns in a

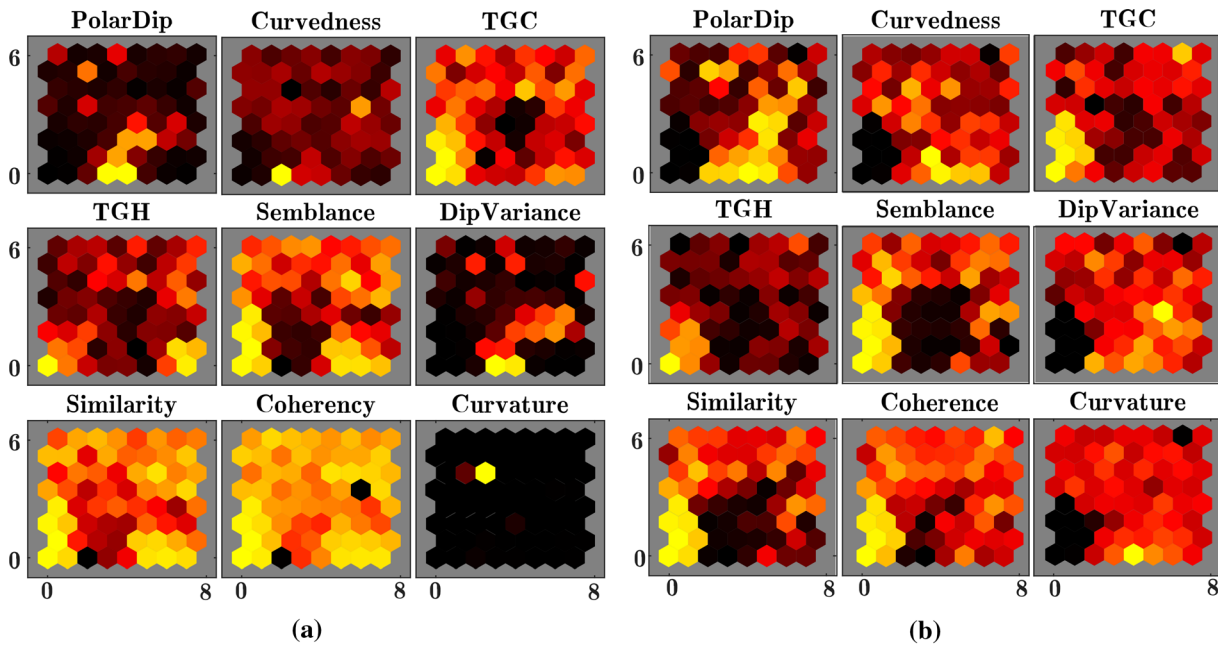**Figure 6.** Unsupervised workflow of training SOM and the datasets and feature selection process applied.

(a)

(b)

Figure 7. Weight plot maps of applying SOM for the feature selection process: (**a**) 2D lines; (**b**) pickset data. These plots show the hexagonal (8 × 8) geometry of the lattice and the weights associated with each of the features. The darker the shade of a hexagon, the larger the weight.

seismic interpretation practice. The detected patterns are justified with geologic data of a study area such as well logs, and the justified patterns can be generalized in a supervised learning scheme. In this

regard, and to show the applicability of unsupervised learning in automatic seismic interpretation workflow and comparing the results with supervised learning, we implemented SOM due to its sensitivity

Table 2. Features with similar weight plots (Fig. 7)

2D lines	Pickset data
Polar dip—Dip variance	**Polar dip**—Dip variance
TGC – TGH	**TGC** – TGH
Similarity – Coherence	**Similarity** – Semblance
Curvedness	**Curvedness**—Curvature
Curvature	**Coherence**
Semblance	

The selected features are in bold

Table 3. LS scores of the nine features

Attributes	Pickset LS	2D Lines LS
Polar dip	0.984	0.992
Curvedness	0.968	0.539
TGC	0.985	0.992
TGH	0.979	0.996
Semblance	0.987	0.996
DipVariance	0.976	0.984
Similarity	**0.989**	0.984
Coherence	0.974	0.982
Curvature	0.950	**1.000**

The highest scores are in bold

to geometries in a dataset to identify clusters in our data.

SOM is a clustering neural network algorithm introduced by Kohonen (2001). SOM is an arrangement of neurons placed on nodes of a 2D lattice ($n \times m$), which is a low dimension grid map connecting the neurons. Each of the neurons contains a vector of initial weights (which are set to random values) and has the same dimension as the input dataset (Roy et al., 2010). The lattice overlays every feature vector (attributes) and changes its geometry to match its structure to the input vector aiming at separating clusters of the data as much as possible. The neurons compete to match the weights to the feature vector, and the winner neuron is the one with weights that best match to it and the other competitive neurons cooperate with the winner and adjust the lattice structure for better definition of the centers of clusters (Ramirez et al., 2012). As a result, weights are allocated to every feature vectors according to their separation capabilities and a feature with larger weights is more capable of separating the clusters.

Supervised Learning

In this study, we used two capable learning devices, SVM and MLP neural networks, to exploit the ability of the selected features in salt detection. We designed them through MATLAB software. A tenfold cross-validation was used to choose the best performance by the contribution of the whole randomly selected data set in training the network (Priddy & Keller, 2005; Van Der Heijden et al., 2005).

Kernel-based SVM methods are wildly popular in pattern classifications and the training process of a linearly inseparable data set. The m-dimensional (here four-dimensional, one for each selected attribute) input space (x) changes to l-dimensional feature space (z). In the z-space, the algorithm can find implicitly the optimum hyperplane by applying radial basis kernel function for delineating classes (Abe, 2010). SVM uses some of the selected vectors in the feature space to obtain the optimal hyperplane, and they are called support vectors (Zhao et al., 2015).

The MLP neural network we used here is a feedforward pattern recognition network that maps the n input neurons (sample data of the four attributes) to the output space. This algorithm uses the whole data set for training as input neurons. Four hidden layers (50, 30, 20, and 10 neurons in each layer) and one output layer are applied to achieve the binary classification output of salt and non-salt. Each neuron implements an activation function in the four hidden and output layers on the data, and this process is optimized with a backpropagation algorithm to attain the lowest classification error. The tangent sigmoid used as activation function for the neurons and resilient backpropagation was the employed train function.

APPLICATION AND WORKFLOWS

In this section, the function of the feature selection procedures and the implementation of the learning devices is presented in unsupervised and supervised workflows.

Unsupervised Workflow

One of the advantages of SOM is that they learn classes and patterns without the need for a
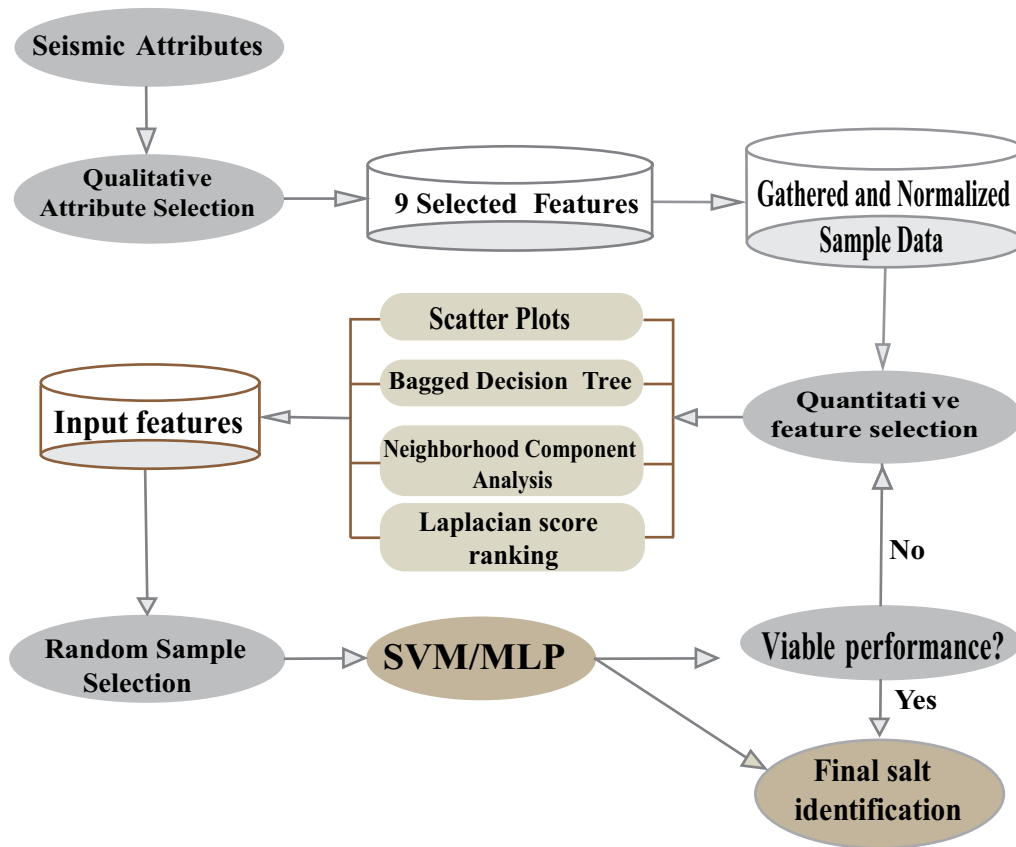
Figure 8. The implemented workflow for supervised salt identification. Qualitative attribute selection followed by four quantitative feature selection is applied before entering the learning process.

labeled picked dataset. Here, two training datasets were used. The first dataset consists of the three randomly selected lines in each direction, namely from the cube in-line 394, line 891, and time slice 1800. The second dataset (pickset data) was selected to test the feasibility of using unlabeled pickset data in the unsupervised method and improving the final results.

A lattice with 8×8 dimension and hexagonal geometry was employed. The lattice size is influential in the number of classes the SOM can detect; thus, it should be assigned with more nodes to save the possibility of seeing unknown classes in the data. The inherent feature weighting of SOM was used in the unsupervised workflow (Fig. 6) to select non-redundant features. Each of the datasets was trained with the nine features, and the weight plot maps were plotted (Fig. 7). These maps display the geometry of the lattice and weights connecting the inputs and the neurons. Because redundant features

represent similar weight plots (Table 2) and because just one of them is enough for the training process, the results of LS (Table 3) were used to complement these weights and pick the features with matching weights to be involved in classification. The selected attributes are displayed in bold in Table 2. After finalizing the training features, the 2D lines and the pickset data input in the training process resulted in two SOM networks with capability of mapping clusters in a new dataset with similar input vectors.

Supervised Workflow

Based on the feature selection results, four attributes (dip variance, similarity, curvature, and TGC) were selected and entered into the learning process. Although Qi et al. (2020) denoted that curvature does not perform well in salt detection, bagged decision trees and NCA prioritized it for the
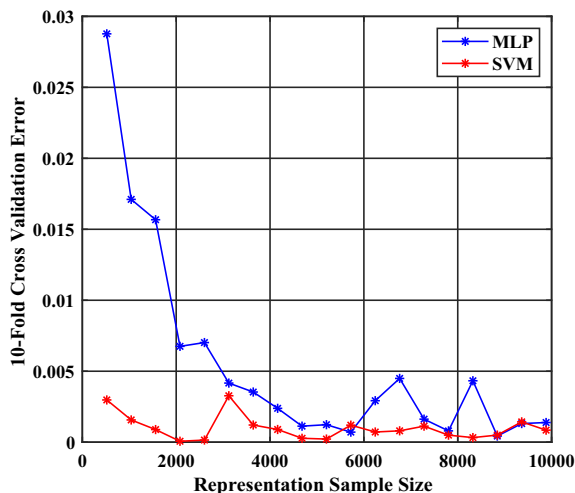
Figure 9. Learning curves showing MSE by stepwise 5% increase in the representative sample size. On the one hand, SVM showed low fluctuation growing the population, and it changed around 0.0025. On the other hand, the training data population affected MLP and its error decreased more than 10 times at the end of the graph.

Table 4. Confusion matrix for the application of the designed SVM to the test data

SVM	Total accuracy: 99.91%			Prediction	
	MSE: 0.0008			Salt	Non-salt
	FP rate: 0.00029	Actual	Salt	3308	1
	F1 score:0.0657		Non-salt	5	3447
MLP	Total accuracy: 99.79%			Prediction	
	MSE: 0.002			Salt	Non-salt
	FP rate: 0.00145	Actual	Salt	3304	5
	F1 score: 0.0656		Non-salt	9	3443

Table 5. Training and test datasets run time by SVM/MLP networks

Parameter	SVM	MLP
Train time (sec)	137.17	25.74
Train data population	3641	3641
Test time (sec)	0.0162	0.035
Test data population	6761	6761

representation sample size in each step (Fig. 9). It shows the influence of the data set population in enhancing the performance of MLP and the limited error fluctuation of SVM due to the classification algorithm based on support vectors. These plots are useful selecting a training set with minimum possible training time and computational costs.

For a fair comparison of training time and accuracy results of the SVM/MLP trained networks, we chose the lowest data population compatible with both of them, and 35% of the pickset data was applied for training the networks. The remaining samples of the data set were used for testing the networks and for creating the confusion matrix. The designed SVM and MLP networks showed high-performance based on the training test data set, with 99.9% accuracy for SVM and 99.79% for MLP (Table 4).

The run time is presented in Table 5 to compare the computational costs of training and testing the networks. Although the MLP training time was about one-fifth of SVM with the same dataset, the test data set classified with SVM used half of the MLP test time. Besides, SVM had a more straightforward network design compared to MLP. The promising results (total accuracy and test time) of SVM were achieved at a higher computational cost (Zhao et al., 2015).

RESULTS

In this section, the results of applying the trained networks based on the selected features are presented. Two seismic lines, including cross-line 1200 and time slice 1832 from the cube, were used to display the classification results. The highest probability of salt presence was annotated on the seismic lines (Fig. 10). The Zechstein salt in these sections does not show a smoothed and homogeneous weak reflection. There were deviated reflectors and heterogeneities that influenced the smoothness of the outputs of the attributes. As displayed in Figure 10a, the marginal layers over the salt dome with distorted amplitudes and the weak continuity of amplitude at the left side are sources of artifacts and can affect facies segmentation. Samples of the selected lines were not included in the training process, and the features were also normalized in the range 0–1.

The first method applied to the lines was the SOM. This clustering algorithm was trained with two
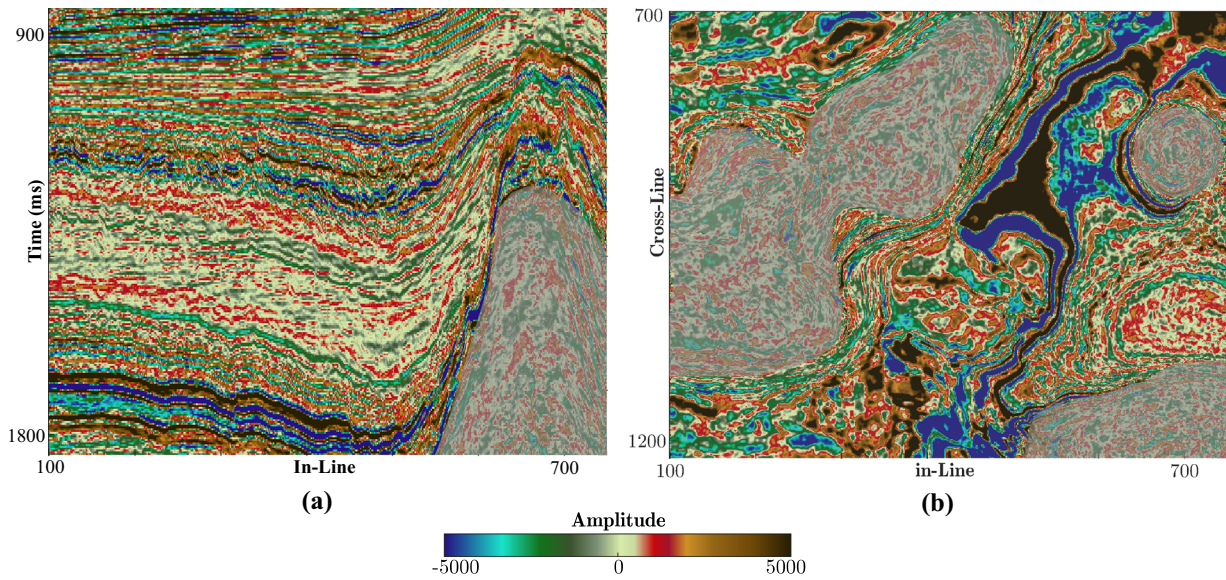
salt classification. The classification was continued by designing SVM/MLP devices, which were applied to the selected seismic lines (Fig. 8).

Learning curves were plotted to assess the two learning algorithms' performances with increasing the data set population. The plot shows MSE by a five percent increase in the tenfold cross-validated

Figure 10. Seismic lines used for deploying the methods and demonstrating the results: (**a**) cross-line 1200; (**b**) 1832 time slice. The locations with highest probability of salt exposure are annotated and shaded (modified from Alaudah et al., 2019).

training data, three seismic lines (Fig. 11a, b), the pickset data (Fig. 11c, d), and their specific feature set to see if a picked dataset can improve SOM results. The SOM classifier detects clusters in a dataset without defining the class or assigning exact probabilities. We normalized the class indices to the range of 0–1 and mapped the results (Fig. 11). Similar ranges in these plots characterize data clusters, and an analogous colormap with SVM/MLP results was allocated for comparison.

The dominant classes in the selected attributes were salt and non-salt, and the goal in the classification was to segment the data to these classes and generalize the patterns to the entire cube. Here, the SOM trained by the selected seismic lines could not segregate the salt in the data and only a general pattern was extracted. The same network trained by a collected unlabeled sample of the two classes resulted in noticeably improved results in the same seismic lines. Although the background was clustered to two classes, and this can complicate the facies detection problem, the salt structures were differentiated and the presence of a prominent facies was revealed. The cross-line and time slice were classified by the trained, supervised networks, and the assigned salt probabilities are plotted in Figure 12. The MLP evaluated every sample; however, SVM used class boundary samples (support vectors) to classify the data, and this resulted in fewer false-

positive (FP) samples in results of the latter compared to results of the former. These are consistent with the confusion matrix (Table 4). In MLP sections, the results were smooth salt exposures (Fig. 12a, b), whereas there was better precision in the classes segmented by SVM (Fig. 12c, d). In regions like salt boundaries, the smoothness of MLP made it hard to follow exact salt edges. In contrast, although SVM performed better in the boundaries it resulted in salt-and-pepper pattern (Fig. 12c).

The seismic lines (in-line 233, cross-line 891, 1836 time slice) and the results of the applied methods in 3D are presented in Figure 13. These plots show the applicability of a multi-attribute method in segmenting salt with any geometry in all directions in a cube. The results of the SOM network trained by the pickset data (Fig. 13a) and the resulting clusters are overlaid on the seismic data with the same colormap as SVM/MLP. The SOM did not cluster the data to two binary classes, and so the overlaid clusters in the salt and non-salt regions included middle colors. The MLP sections (Fig. 13b) show smooth results, whereas the SVM results (Fig. 13c) display more restricted salt probability regions, the lowest amounts of artifacts, and the sharpest boundaries.
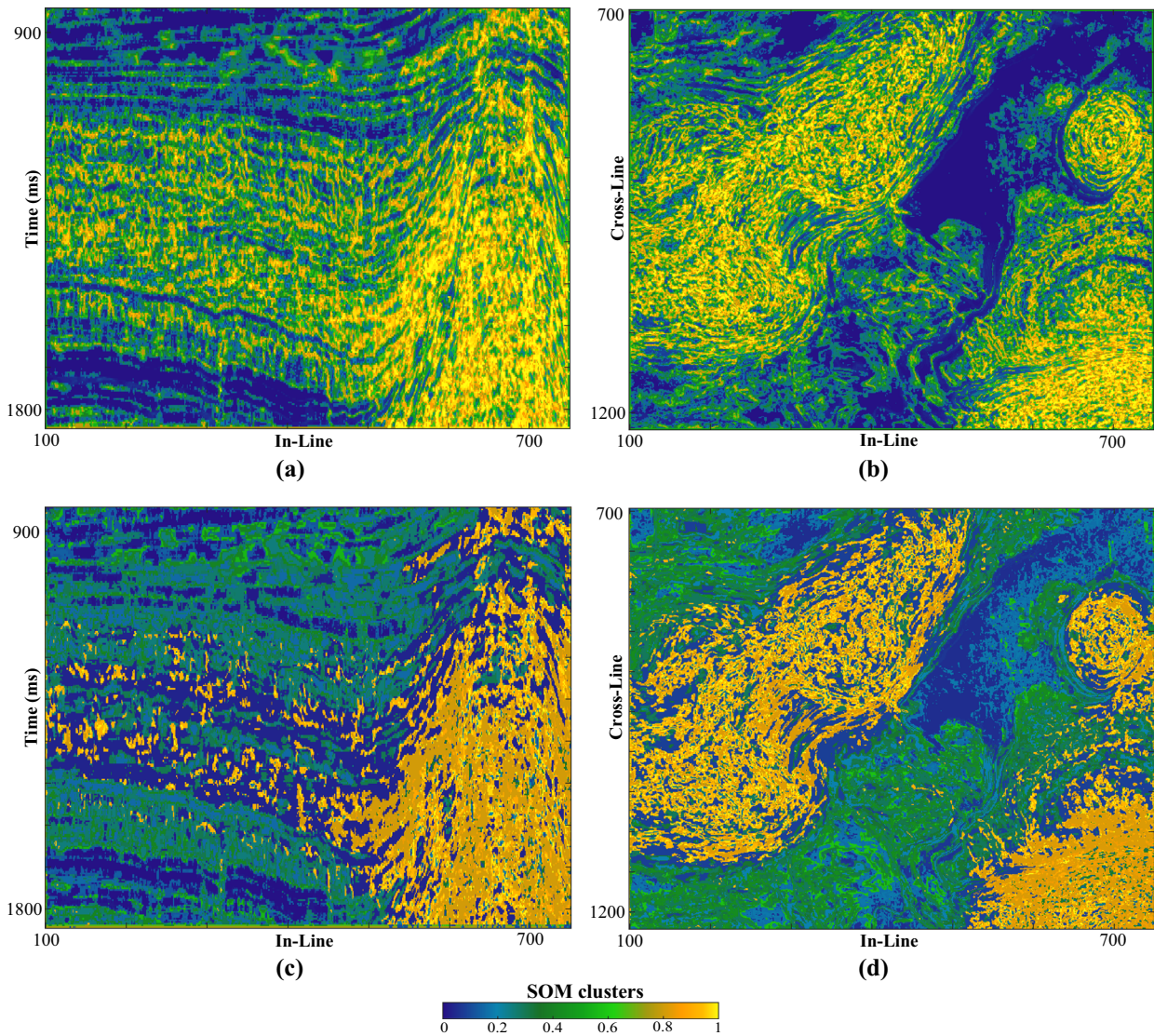
Figure 11. Lines clustered by the SOM: (**a**) and (**c**) cross-line 1200; (**b**) and (**d**) 1832 time slice. Top row: SOM results trained by seismic lines. Bottom row: SOM results trained by unlabeled pickset and the improved classification results.

DISCUSSION

This study employed (a) nine seismic attributes with sensitivity to chaotic features but relatively insensitive to seismic facies boundaries, and (b) SOM, SVM, and MLP techniques. In calculating the attributes, full dip-steered data were used to consider dip changes in the final outputs. Utilizing the attributes of different nature reduces the chances of misclassification of a weak feature (Di et al., 2019) and, considering diverse characteristics, results in an adequately classified output. By using different ML

methods on a common dataset, it is possible to compare the benefits and deficiencies of the methods (Chopra & Marfurt, 2018).

Suitable attributes for interactive interpretation may not work well in ML procedures (Infante-Paez & Marfurt, 2019). Additionally, due to the large scale of seismic volumes, attributes dimensionality reduction was employed to accelerate multi-attributes ML workflows, reduce memory demand (Kim et al., 2019; Roden et al., 2015), prevent misclassification, and raise interpretability of the results.
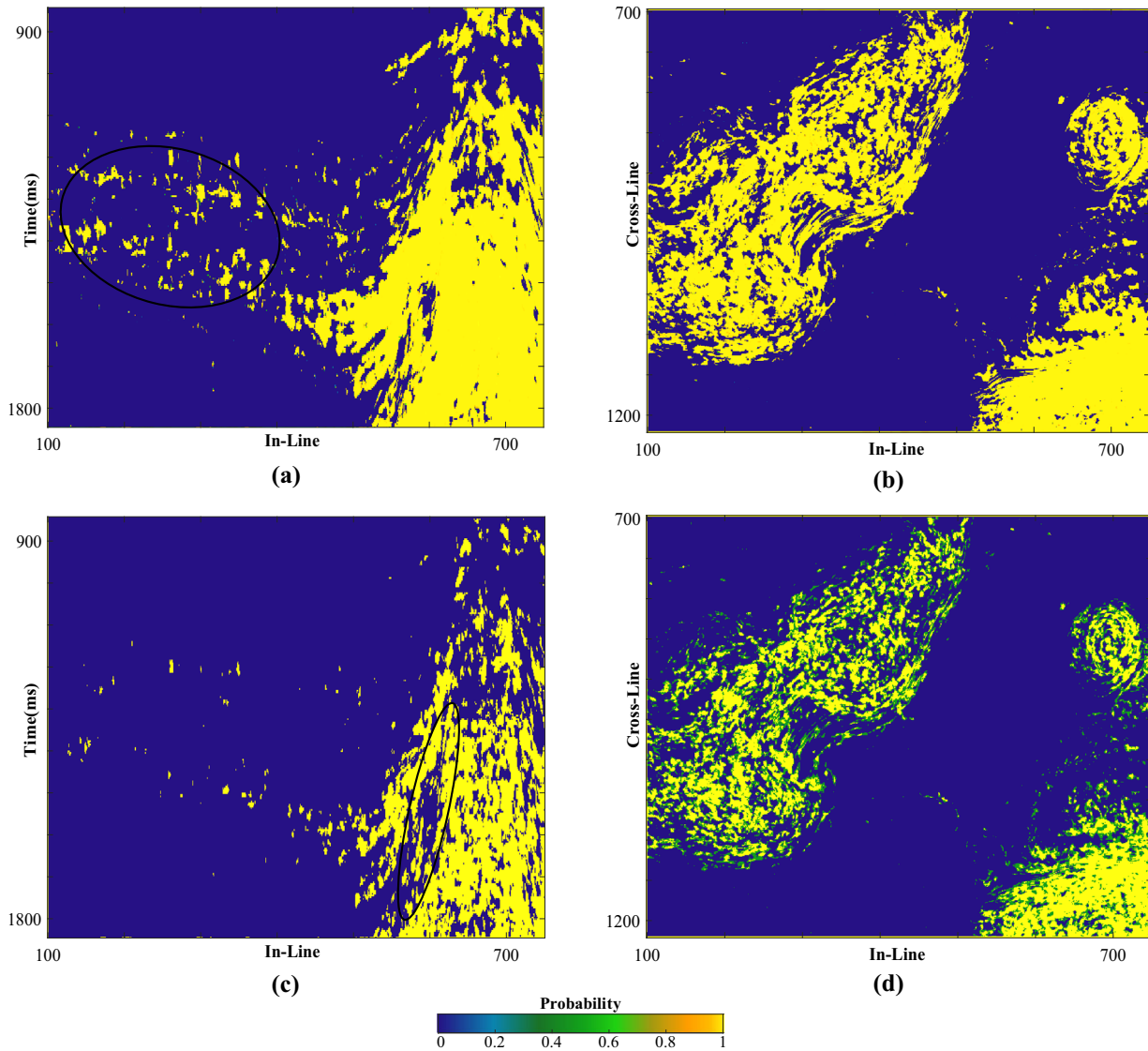
Figure 12. Salt probability assigned by supervised methods with 1 being the highest (salt): (**a**) and (**c**) cross-line 1200; (**b**) and (**d**) 1832 time slice. Top row: MLP. Bottom row: SVM. The ellipsoid area in (**a**) shows the misclassified samples (false positives) at the left side of the salt dome. In (**c**), the differentiated salt dome from the marginal layers shows similar amplitude patterns.

SVM and MLP can efficiently incorporate multi-attribute sets and perform supervised classifications. The MLP with 0.00145 false-positive rates had an acceptable performance on the test dataset and was applied for salt classification (Fig. 13b). The results of MLP are smooth and, in the boundary regions, show some misclassified samples and smoothed patterns for the marginal layers. Despite this, the same effect on the internal salt resulted in mitigation of the salt-and-pepper outputs that were probable by other classifiers. Although SVM reduced the false-positive rate to 0.00029, the salt-and-

pepper view reduced the smoothness of internal salt structures (Fig. 13c). However, along the salt border, the exact probability allocation property of SVM detected more patterns. By projecting the data into a higher-dimensional space, SVM builds a linearly separable dataset. This causes the increased accuracy, obtained at the cost of increased computational cost (Zhao et al., 2015). In a small training dataset, there is a possibility of over-fitting MLP and reducing generalization quality. Thus, this network must be trained with proper data samples and tested accurately to assess its quality. MLP results are not

(a)

(b)

(c)
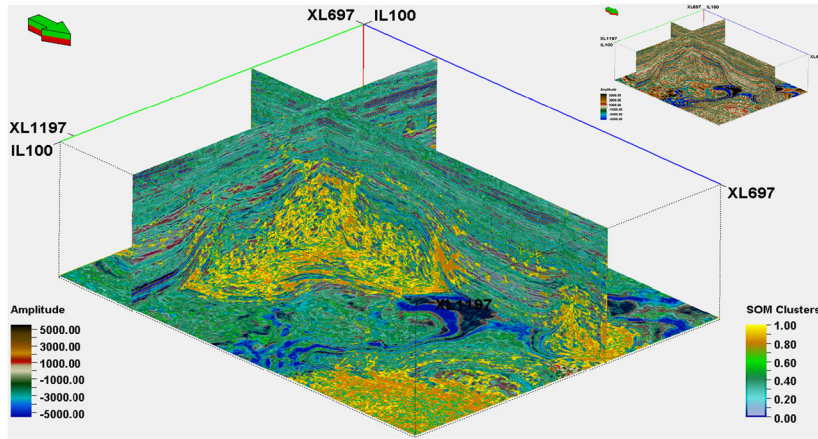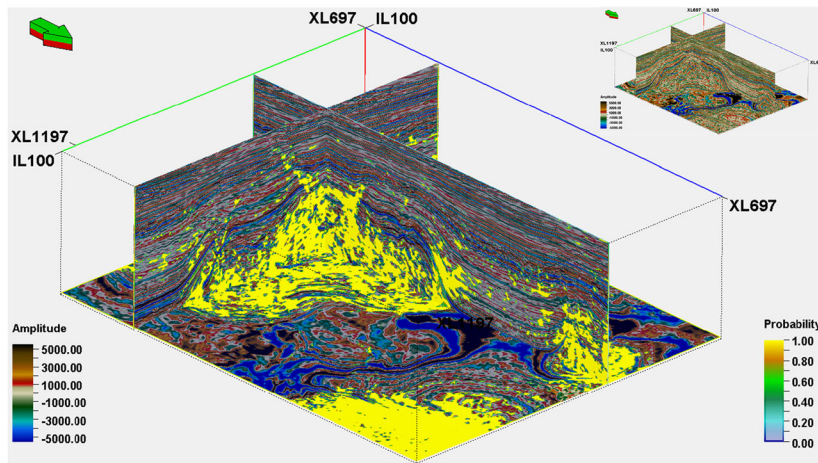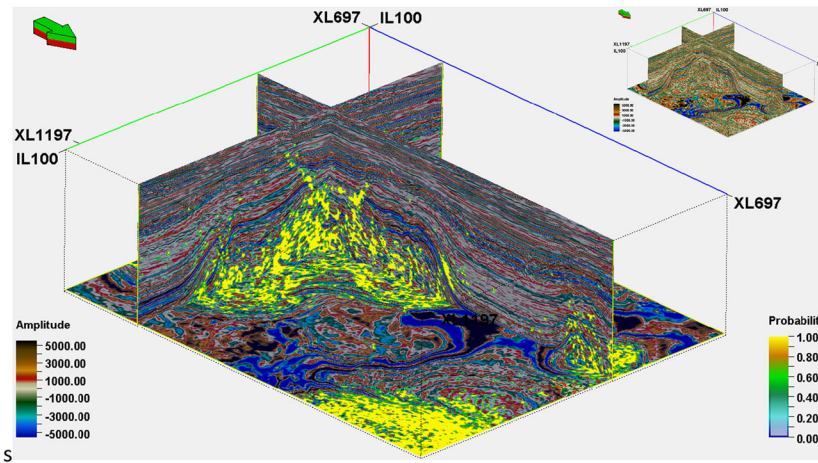
◄**Figure 13.** 3D views of salt exposures in F3 block including in-line 233, cross-line 891, 1836 time slice: (**a**) SOM trained by pickset data; (**b**) SVM result; (**c**) MLP result. Inset in every plot shows the underlying seismic data. The background colormap is the seismic amplitude, and an overlay of the classes is plotted with similar colormap for comparison; along with transparency are colors representing less than 0.5 figures.

stationary, and every training with the same dataset changes the weights, confusion matrix, and network performance. The SVM does not over-fit and, with a deterministic algorithm, results in an invariable network (Bagheri & Riahi, 2015).

The extraction of the chaotic feature is complicated by complex subsurface geology, intrusion of salt through different tectonic processes, heterogeneities of salt texture, and distorted and highly dipping reflectors on its flanks. Thus, attributes may fail in differentiating samples precisely and presenting a solid presence of a chaotic structure, and misclassified samples are observed in the sections. The SVM classifier can distinguish these samples and assign sharper probabilities, but at the expense of showing crisper results.

Overall, the salt trace could be followed in all of the sections in the cube (in-line, cross-line, time slice) with one step of training (Fig. 13) due to the geometry of the picked samples. In these sections, the results are in good agreement with the salt distribution of the F3 Block presented by different methods (Shafiq et al., 2017; Di et al., 2018; Alaudah et al., 2019). The in-lines and cross-lines display concentrated salt distribution. They are slightly different from the highest amplitude contrast, which can help pick the salt with any geometry or indefinite boundary.

Some studies focused on improving the processing step (Alaei et al., 2018; Farrokhnia et al., 2018) or filtering methods to enhance salt segmentation (Qi et al., 2016). As asserted by Qi et al. (2016), using smoothing filters on the attributes helps to define chaotic features in a scale that can be seen by interpreters and solve the salt-and-pepper view of the target facies. Although this pre-conditioning increases the smoothness of the results for the SVM classifier, this filtering of attributes raises MLP and SOM misclassification errors. Thus, in this paper, we presented the results without smoothing the attributes to compare the classifiers' potential in reducing the salt-and-pepper observations.

Because subsurface geology is local, variant and the seismic surveys have different acquisition parameters, it is challenging to use the same attributes and repeat an interpretation workflow. However, these workflows can be repeated with various datasets or new features to enhance the detection of chaotic facies.

CONCLUSION

Supervised learning teaches a machine to recognize the schemes introduced by the interpreter. It is helpful for interpreting subtle patterns and for certifying and accelerating the results of analyzing big data sets. The multi-attribute approach helps the classifier differentiate classes, and using attributes of different natures (e.g., geometric, texture, and edge detector attributes) is a feasible approach. Whenever an attribute is not responding, the classifier can refer to other features in each sample. This study employed three supervised and unsupervised techniques to elevate the realization of details in seismic sections. The results showed that the performance of SOM is noticeably enhanced by using an unlabeled dataset, and it could detect all the classes of the data with higher precision. MLP with shorter training time and good accuracy resulted in more smoothed inner salt results. This makes it even a good choice for seismic features with continuity as one of their characteristics (faults and channels). Along boundaries, SVM showed more distinct classes and reduced the misclassification error to more than two times than that of MLP but with higher computational costs. Overall, none of the implemented methods outperformed the others in every aspect, and each of them has its own benefits and usage. The balance between the network selection and the classification or clustering goal is vital. Computational cost, target feature, feature selection algorithms, and network design are to be considered in every ML application. Sampling and processing 2D sections in 3D helps overcome computational resource limitations, making it possible to follow a facies distribution in a cube. Furthermore, the workflows can fit any facies detection task and can be updated by adding more representative attributes.

ACKNOWLEDGMENTS

REFERENCES

Abe, S. (2010). Feature selection and extraction. In *Support vector machines for pattern classification* (pp. 331–341). London, UK: Springer.

Alaei, N., Roshandel Kahoo, A., Kamkar Rouhani, A., & Soleimani, M. (2018). Seismic resolution enhancement using scale transform in the time-frequency domain. *Geophysics, 83*(6), V305–V314.

Alaudah, Y., Michałowicz, P., Alfarraj, M., & AlRegib, G. (2019). A machine-learning benchmark for facies classification. *Interpretation, 7*(3), SE175–SE187.

Bagheri, M., & Riahi, M. A. (2015). Seismic facies analysis from well logs based on supervised classification scheme with different machine learning techniques. *Arabian Journal of Geosciences, 8*(9), 7153–7161.

Bahorich, M., & Farmer, S. (1995). 3-D seismic discontinuity for faults and stratigraphic features: The coherence cube. *The Leading Edge, 14*(10), 1053–1058.

Barnes, A. E. (2016). *Handbook of poststack seismic attributes*. Society of Exploration Geophysicists.

Berthelot, A., Solberg, A. H., & Gelius, L. J. (2013). Texture attributes for detection of salt. *Journal of Applied Geophysics, 88*, 52–69.

Breiman, L. (2001). Decision-tree forests. *Machine Learning, 45*(1), 5–32.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. CRC Press.

Buist, C., Bedle, H., Rine, M., & Pigott, J. (2021). Enhancing Paleoreef reservoir characterization through machine learning and multi-attribute seismic analysis: silurian reef examples from the Michigan Basin. *Geosciences, 11*(3), 142.

Chopra, S., & Marfurt, K. J. (2018). Seismic facies classification using some unsupervised machine-learning methods. In *2018 SEG international exposition and annual meeting*. OnePetro. https://doi.org/10.1190/segam2018-2997356.1.

Chopra, S., & Alexeev, V. (2006). Applications of texture attribute analysis to 3D seismic data. *The Leading Edge, 25*(8), 934–940.

Chopra, S., & Marfurt, K. J. (2007). *Seismic attributes for prospect identification and reservoir characterization*. Society of Exploration Geophysicists and European Association of Geoscientists and Engineers.

Di, H., & AlRegib, G. (2017). Seismic multi-attribute classification for salt boundary detection—A comparison. Paper presented at the *79th EAGE conference and exhibition 2017*.

Di, H., & AlRegib, G. (2020). A comparison of seismic salt body interpretation via neural networks at sample and pattern levels. *Geophysical Prospecting, 68*(2), 521–535.

Di, H., Shafiq, M., & AlRegib, G. (2018). Multi-attribute k-means clustering for salt-boundary delineation from three-dimensional seismic data. *Geophysical Journal International, 215*(3), 1999–2007.

Di, H., Shafiq, M. A., Wang, Z., & AlRegib, G. (2019). Improving seismic fault detection by super-attribute-based classification. *Interpretation, 7*(3), SE251–SE267.

Eichkitz, C. G., Schreilechner, M. G., de Groot, P., & Amtmann, J. (2015). Mapping directional variations in seismic character using gray-level co-occurrence matrix-based attributes. *Interpretation, 3*(1), T13–T23.

Farrokhnia, F., Kahoo, A. R., & Soleimani, M. (2018). Automatic salt dome detection in seismic data by combination of attribute analysis on CRS images and IGU map delineation. *Journal of Applied Geophysics, 159*, 395–407.

Fossen, H. (2016). *Structural geology*. Cambridge University Press.

Gao, D. (2007). Application of three-dimensional seismic texture analysis with special reference to deep-marine facies discrimination and interpretation: Offshore Angola, west Africa. *AAPG Bulletin, 91*(12), 1665–1683.

Hall-Beyer, M. (2000). GLCM texture: A tutorial. *National Council on Geographic Information and Analysis Remote Sensing Core Curriculum, 3*, 75.

He, X., Cai, D., & Niyogi, P. (2005). Laplacian score for feature selection. *Advances in Neural Information Processing Systems, 18*.

Infante-Paez, L., & Marfurt, K. J. (2019). Using machine learning as an aid to seismic geomorphology, which attributes are the best input? *Interpretation, 7*(3), SE1–SE18.

Kim, Y., Hardisty, R., & Marfurt, K. J. (2019). Attribute selection in seismic facies classification: Application to a Gulf of Mexico 3D seismic survey and the Barnett Shale. *Interpretation, 7*(3), SE281–SE297.

Koenderink, J. J., & Van Doorn, A. J. (1992). Surface shape and curvature scales. *Image and Vision Computing, 10*(8), 557–564.

Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Berlin: Springer. https://doi.org/10.1007/978-3-642-56927-2.

La Marca-Molina, K., Silver, C., Bedle, H., & Slatt, R. (2019). Seismic facies identification in a deepwater channel complex applying seismic attributes and unsupervised machine learning techniques. A case study in the Taranaki Basin, New Zealand. *SEG Technical Program*. https://doi.org/10.1190/segam2019-3216705.1.

Marfurt, K. J. (2006). Robust estimates of 3D reflector dip and azimuth. *Geophysics, 71*(4), P29–P40.

Marfurt, K. J., Kirlin, R. L., Farmer, S. L., & Bahorich, M. S. (1998). 3-D seismic attributes using a semblance-based coherency algorithm. *Geophysics, 63*(4), 1150–1165.

Odoh, B. I., Ilechukwu, J. N., & Okoli, N. I. (2014). The use of seismic attributes to enhance fault interpretation of OT Field, Niger Delta. *International Journal of Geosciences, 5*, 826–834.

Priddy, K. L., & Keller, P. E. (2005). *Artificial neural networks: An introduction* (Vol. 68). SPIE Press.

Qi, J., Lin, T., Zhao, T., Li, F., & Marfurt, K. (2016). Semisupervised multiattribute seismic facies analysis. *Interpretation, 4*(1), SB91–SB106.

Qi, J., Zhang, B., Lyu, B., & Marfurt, K. (2020). Seismic attribute selection for machine-learning-based facies analysis. *Geophysics, 85*(2), O17–O35.

Ramirez, C., Argaez, M., Guillen, P., & Gonzalez, G. (2012). Self-organizing maps in seismic image segmentation. *Computer Technology and Application, 3*(9), 624–629.

Remmelts, G. (1996). *Salt tectonics in the southern North Sea, the Netherlands* (pp. 143–158). Springer Netherlands. https://doi.org/10.1007/978-94-009-0121-6_13.

Roberts, A. (2001). Curvature attributes and their application to 3D interpreted horizons. *First Break, 19*, 85–100.

Roden, R., Smith, T., & Sacrey, D. (2015). Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps. *Interpretation, 3*(4), SAE59–SAE83.

Rounds, E. (1980). A combined nonparametric approach to feature selection and binary decision tree design. *Pattern Recognition, 12*(5), 313–317.

Roy, A., Matos, M., & Marfurt, K. J. (2010). Automatic seismic facies classification with Kohonen self-organizing maps—A tutorial. *Geohorizons Journal of Society of Petroleum Geophysicists, 15*, 6–14.

Shafiq, M. A., Wang, Z., AlRegib, G., Amin, A., & Deriche, M. (2017). A texture-based interpretation workflow with application to delineating salt domes. *Interpretation, 5*(3), S1–S19.

Soleimani, M., Aghajani, H., & Heydari-Nejad, S. (2018). Salt dome boundary detection in seismic image via resolution enhancement by the improved NFG method. *Acta Geodaetica Et Geophysica, 53*(3), 463–478.

Tingdahl, K. M., Bril, B., & de Groot, P. (2002). Simultaneous mapping of faults and horizons with the help of object probability cubes and dip-steering. In *SEG Technical Program Expanded Abstracts 2002* (pp. 520–523). Society of Exploration Geophysicists.

Tingdahl, K. M. (2003). Improving seismic chimney detection using directional attributes. *Developments in Petroleum Science, 51*, 157–173. https://doi.org/10.1016/S0376-7361(03)80013-4.

Tingdahl, K. M., Bril, A. H., & de Groot, P. F. (2001). Improving seismic chimney detection using directional attributes. *Journal of Petroleum Science and Engineering, 29*(3–4), 205–211.

Tingdahl, K. M., & De Rooij, M. (2005). Semi-automatic detection of faults in 3D seismic data. *Geophysical Prospecting, 53*(4), 533–542.

Van Der Heijden, F., Duin, R. P., De Ridder, D., & Tax, D. M. (2005). *Classification, parameter estimation and state estimation: An engineering approach using MATLAB*. Wiley.

Waldeland, A., & Solberg, A. (2017). Salt classification using deep learning. Paper presented at the *79th EAGE conference and exhibition 2017*.

Yang, W., Wang, K., & Zuo, W. (2012). Neighborhood component feature selection for high-dimensional data. *Journal of Computers, 7*(1), 161–168.

Yenugu, M., Marfurt, K. J., & Matson, S. (2010). Seismic texture analysis for reservoir prediction and characterization. *The Leading Edge, 29*(9), 1116–1121.

Zhao, T., Jayaram, V., Roy, A., & Marfurt, K. J. (2015). A comparison of classification techniques for seismic facies recognition. *Interpretation, 3*(4), SAE29–SAE58.