

Seismic structure interpretation based on machine learning: A case study in coal mining

Dong Li¹, Suping Peng², Yongxu Lu¹, Yinling Guo¹, and Xiaoqin Cui²

Abstract

Interpretation of geologic structures entails ambiguity and uncertainties. It usually requires interpreter judgment and is time consuming. Deep exploitation of resources challenges the accuracy and efficiency of geologic structure interpretation. The application of machine-learning algorithms to seismic interpretation can effectively solve these problems. We analyzed the theory and applicability of five machine-learning algorithms. Seismic forward modeling is a key connection between the model and seismic response, and it can obtain seismic data of known geologic structures. Based on the modeling data, we first optimized the seismic attributes sensitive to the target geologic structure and then we verified the accuracy of the five machine-learning algorithms by the cross-checking method. In this case, the random forest algorithm had the highest accuracy. So we examined the structural interpretation method based on a random forest using the 3D seismic reflection data from coalfield exploration. The prediction effect of this interpretation workflow is verified by comparison with known geologic structures on the plane and profile. The results suggest that the random forest algorithm is feasible to indicate geologic structure interpretations in the case of collapsed column and fault structures and it can effectively improve the efficiency of seismic interpretation and its accuracy. The machine-learning-based workflow provides a new technique for seismic structure interpretation in coal mining.

Introduction

Seismic data contain a variety of information in a multidimensional setting, including geometry, kinematics, dynamics, and statistical features. They are often described by seismic attributes. The concept of the seismic attribute originated in the 1960s, and it solved prediction problems of stratigraphic structure, lithology, and physical properties to varying degrees.

More than 100 kinds of seismic attributes have been proposed for different prediction targets. Seismic attributes have been used to detect buried river channels (Liu and Marfurt, 2007) and buried karsts (Jamaludin et al., 2017). These attributes are also widely used in detecting gas and hydrate gas associations (Lee et al., 2009), discontinuity in sedimentary layers, disconformities, lithology, sedimentary environments changes, and reservoir characteristics (Chopra and Marfurt, 2005, 2008, 2009, 2012; Hosseini et al., 2011; Anees, 2013; Dezfoolian et al., 2013; Koson et al., 2014; Maleki et al., 2015; Oyeyemi and Aizebeokhai, 2015; Shamsuddin et al., 2017).

Seismic interpretation is complex and usually requires comprehensive use of various seismic attributes. The

geologic structure interpretation method based on seismic attributes has been the traditional means in the past few decades. However, this method only uses one or a few attributes, which increases the risk of interpretation errors and requires subjective judgment and interpretation experience. Later, some researchers have proposed a semiautomatic interpretation method, which improved the work efficiency to some extent (Wang and AlRegib, 2017). With the development of computers, the emergence of machine learning provides more tools for seismic interpretation. It can integrate multiple seismic attributes into interpretation and has high precision and efficiency. Several common machine-learning algorithms exist, such as decision trees, random forests, neural networks, support vector machines (SVMs), the genetic algorithm (GA), and naive Bayes classifiers. In the field of resource exploration and development, machine-learning algorithms are used for lithology facies analysis (Zhao et al., 2014; Wrona et al., 2018) and fault and salt-dome detection (Zhang et al., 2014; AlRegib et al., 2018; Wang et al., 2018b). At the same time, the recent emerging convolutional neural network (CNN) has also been applied to

¹China University of Mining and Technology (Beijing), State Key Lab of Coal Resources and Safe Mining, Beijing 100083, China and China University of Mining and Technology (Beijing), College of Geoscience and Surveying Engineering, Beijing 100083, China. E-mail: lidong_6666@163.com (corresponding author); samhomes@163.com; 597965973@qq.com.

²China University of Mining and Technology (Beijing), State Key Lab of Coal Resources and Safe Mining, Beijing 100083, China. E-mail: psp@cumtb.edu.cn; cuixq65@hotmail.com.

Manuscript received by the Editor 9 November 2018; revised manuscript received 1 March 2019; published ahead of production 2 April 2019; published online 28 May 2019. This paper appears in *Interpretation*, Vol. 7, No. 3 (August 2019); p. SE69–SE79, 14 FIGS., 3 TABLES.

<http://dx.doi.org/10.1190/INT-2018-0208.1>. © 2019 Society of Exploration Geophysicists and American Association of Petroleum Geologists. All rights reserved.

the interpretation (Di et al., 2018; Waldeland et al., 2018). It only requires primitive seismic reflection and does not need to extract seismic attributes.

This study uses traditional machine-learning algorithms and is mainly concerned with the identification of collapsed columns and fault structures. The seismic attributes sensitive to the structure are selected and used as the input layer data in the prediction network model to realize the prediction target of the geologic structure.

The paper is divided into three parts. First, machine-learning algorithms and seismic attributes are introduced. Second, we explain how forward modeling synthetic data can be used for the optimization of seismic attributes and discuss different machine-learning algorithms. Finally, we use real seismic data with an appropriate machine-learning algorithm to predict the geologic structures of the study area.

Theory

Machine-learning algorithms

Machine learning belongs to the category of artificial intelligence and statistics. It has been widely used in various industries in recent years. It is an effective means to classify large-scale data and realize intelligent learning. Given the wide application of machine learning, many intelligent algorithms have emerged, which have their own advantages and disadvantages depending on different learning objectives. This paper primarily focuses on the back propagation (BP) neural network, radial basis function (RBF) neural network, SVM, decision tree, and random forest, and it applies them to the detection of geologic structures.

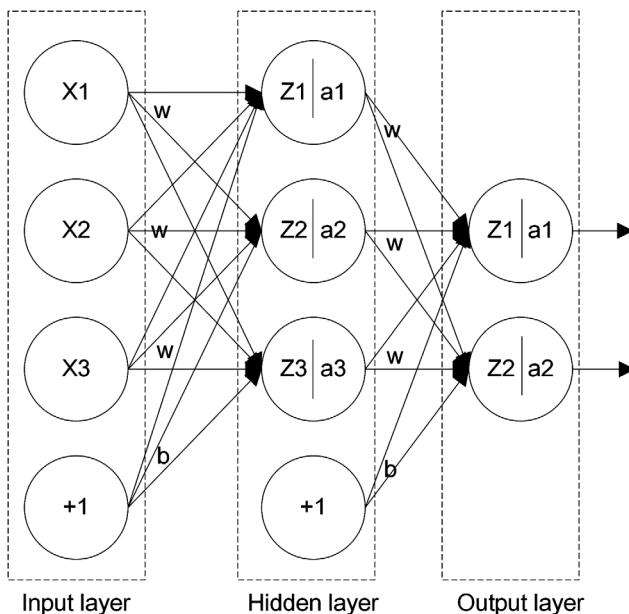


Figure 1. Neural network model.

Artificial neural networks

Artificial neural networks are a series of learning models inspired by biological neural networks that connect a large number of neuron nodes (Figure 1). The neural network model can be divided into a feedforward network and a feedback network. The feedforward network is a neuron that accepts input data from the upper level and transmits them to the next level without feedback. The feedback network is a feedback process between neurons. The paper applies the BP (feedback) and RBF (feedforward) neural networks.

The neural network contains the input layer, hidden layer, and output layer. After the neurons receive input from the previous layer, the input data are weighted and summed, and an offset is added as the input value (Huang et al., 2018). The neuron transforms the input data through the activation function to obtain the output value. The formula is as follows:

$$Z_j^l = \sum_{i=1 \dots n} w_{ij} \cdot a_{ij}^{l-1} + b_j^l, \quad (1)$$

where Z_j^l represents the input value of the j th neuron of layer 1, w_{ij} refers to the connection from the i th neuron to the j th neuron, w represents the weight, a^{l-1} denotes the output value of layer $l-1$, that is, the input value of layer l , and b denotes an offset

$$a_j^l = f(z_j^{(l)}), \quad (2)$$

where a_j^l represents the output value of the j th neuron of layer 1 and f denotes an activation function.

After inputting the data, the BP neural network initializes the weight and offset of each transmission line and performs forward transmission to obtain the predicted value. When the prediction error is greater than the threshold, feedback is performed to update the weight and offset, and forward transmission is performed again. This process is repeated until the error meets the requirements, and the error accuracy is set to 0.01.

The RBF neural network is a kind of nonlinear mapping that can be regarded as mapping the (low-dimension) input layer space to the (high-dimension) hidden layer space and linear mapping of the hidden layer space to the output layer. Classification from low to high dimensions is achieved by selecting the appropriate RBF function.

SVM

SVM is a supervised learning model based on statistical learning theory. It differs from neural networks in that SVM uses mathematical methods and optimization techniques (Rostami et al., 2018; Routray et al., 2018). SVM maps the complex vectors of low-dimensional space into a higher dimensional space, and it constructs the maximum interval hyperplane in this high-dimensional space to classify the data (Figure 2).

For a training sample data set (x, y) , where $\{x_i | i=1, 2, \dots, n\}$, is the input sample, and $\{y_i | i=1, 2, \dots, n\}$

is the output sample. The nonlinear mapping function φ is used to map the data in the low-dimensional space to the high-dimensional space, and then the optimal decision function is constructed in the high-dimensional space as follows:

$$y = \mathbf{w}^T \cdot \varphi(x) + b, \quad (3)$$

where \mathbf{w} is the weight vector, b is the offset, and φ is the mapping function.

The optimization problem of the SVM algorithm can be expressed by the principle of minimization (Wang et al., 2015):

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i, \quad (4)$$

$$\text{s.t. } \begin{cases} y_i(\mathbf{w}^T x_i + b) \geq 1 - \xi_i & i = 1, 2, \dots, m. \\ \xi_i \geq 0 \end{cases} \quad (5)$$

In the formula, C is the penalty coefficient and ξ is the relaxation coefficient.

The key to SVM is the selection of kernel functions. Different kernel functions result in different SVM algorithms. The kernel function can effectively solve the complex computation problem of high-dimensional space classification. To correct the error, two variables, namely, the relaxation coefficient and the penalty coefficient, need to be introduced (Liu and Jiao, 2011). The global-optimization ability of the GA is outstanding, and the best parameters can be obtained (Huang et al., 2017). Therefore, to eliminate the error caused by parameter selection, this paper uses GA to optimize the parameters of SVM. The effective combination of the two algorithms is usually called GA-SVM algorithm. In the calculation of this paper, the final penalty coefficient value is 83.3426 and the relaxation coefficient value is 10.5345.

Decision trees and random forests

The decision tree is a supervised learning method (Marjanović et al., 2018), which adopts a tree structure recursively from top to bottom, starting from the root node of the tree. It compares and tests the attribute values on the internal nodes, and then, according to the given attribute value of the instance, it determines the

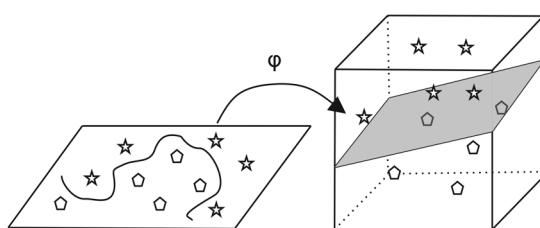


Figure 2. Construction of the hyperplane: φ is the mapping function.

corresponding branch. Finally, the conclusion is obtained at the leaf node of the decision tree (Figure 3).

For the sample data set $T = \{t_1, t_2, \dots, t_m\}$, the information expectation of the sample can be calculated as follows (Wang et al., 2018a):

$$E(T) = - \sum_{i=1}^m p(t_i) \log_2 p(t_i), \quad (6)$$

where m represents the number of categories, and $p(t_i) = |t_i| / \sum_{i=1}^m |t_i|$.

If the sample T is divided according to the attribute A , the calculation formula for obtaining the information gain Gain (T, A) is as follows:

$$\text{Gain}(T, A) = E(T) - E_A(T), \quad (7)$$

where E_A is the information entropy of the sample subset divided by attribute A . If A has k values, the sample T is divided into k subsets $\{T_1, T_2, \dots, T_k\}$, and the formula is as follows:

$$E_A(T) = \sum_{i=1}^k \frac{|T_i|}{|T|} E(T_i). \quad (8)$$

The random forest method was proposed in 2001 (Breiman, 2001). It is an algorithm developed on the basis of decision trees, including classifiers for multiple decision trees, which effectively enhances the generalization ability of decision trees (Dong et al., 2014; Hibert et al., 2017). This ensemble method integrates multiple models and complements each other to solve the defects of a single model and avoid its limitations. The data are resampled by the bagging strategy to generate multiple data sets, thereby establishing multiple decision trees and forming a random forest. The final classification decision formula is as follows:

$$H(x) = \arg \max_Y \sum_{i=1}^k I(h_i(x) = Y), \quad (9)$$

where $H(x)$ is the combined classification model, $h_i(x)$ is the single decision tree classification model, Y is the output variable, and $I(\cdot)$ is the indicator function.

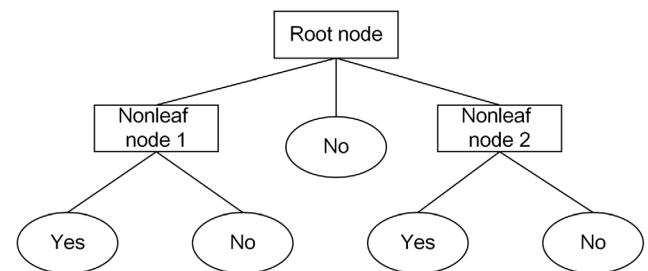


Figure 3. Decision tree structure.

Seismic attribute extraction and analysis

Synthetic data simulation

Forward modeling is an effective way to obtain seismic reflection signals of known geologic structures. The paper analyzes the sensitivity of seismic attributes of target geologic structures using forward modeling data. Therefore, based on the geologic conditions of the study area, including the sedimentary environment and geologic structure development laws, a typical geologic model was first established before the study (Figure 4). It consists of four rock layers, namely, mudstone, coal, mudstone, and limestone, from the top to the bottom. The thickness of the coal seam is approximately 10 m. The geologic parameters of the construction for the collapse column, normal fault, and reverse fault geologic structures are shown in Table 1. In Figure 4, the geologic structures from left to right are collapse column X1, collapse column X2, normal fault F1, and reverse fault

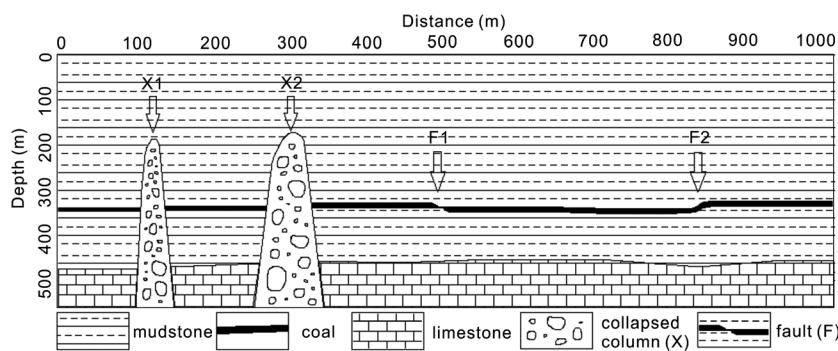


Figure 4. Forward model: X1, X2 are the collapsed column number, and F1, F2 are the fault number.

Table 1. The model's geologic parameters.

Parameter	Stratum			Structure
	Mudstone	Coal	Limestone	
Velocity ($\text{m} \cdot \text{s}^{-1}$)	3000	2500	3600	1800
Density ($\text{kg} \cdot \text{m}^{-3}$)	2200	1500	2700	1100

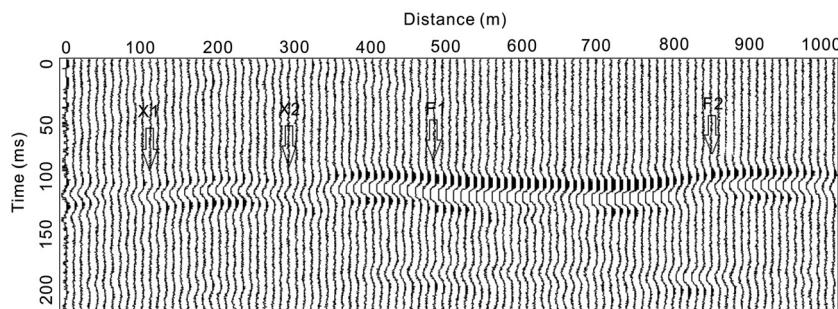


Figure 5. Seismic section: X1, X2 correspond to the position of the collapsed column, and F1, F2 correspond to the fault position.

F2 of the flexural property. The diameters of the long axes of the collapse columns X1 and X2 are 20 and 45 m, respectively. The fault distance of fault F1 is 8 m.

The number of seismic traces is 1000 with the step of 1 m, and the source is a 60 Hz Rayleigh wavelet. Considering the noise factor in the actual data, 5% white noise is added to the model data. The obtained reflected wave time profile is shown in Figure 5. The coal seam reflection is clearly observed on the time profile, and the amplitude of the reflected wave of the coal seam at the collapse column is weakened. However, the change in the reflected wave of the coal seam at the normal fault is small, and the reflected wave of the coal seam at the reverse fault is bent along the same axis.

Seismic attribute optimization

Based on the forward model data, after analyzing all of the seismic attributes provided by Schlumberger Petrel 2015®, 13 properties that can be used to analyze structural discontinuities are selected, as shown in Table 2. They are 3D curvature, dip illumination, local structural dip, variance, amplitude contrast, chaos, cosine of phase, dominant frequency, first derivative, instantaneous frequency, instantaneous quality, local flatness, and second derivative. Some attributes have an obvious response at the construction position, and some attributes have no obvious changes. Figure 6 shows all sensitive attributes and one insensitive attribute extracted along the coal seam of the forward model. It can be seen that the local structural dip

attribute has no characteristic change at the construction position; thus, it is not suitable for identification of this geologic formation. The other five properties show maximum or minimum values at the structure locations. Therefore, these five properties, namely, instantaneous frequency, dip illumination, variance, chaos, and local flatness, are preferred.

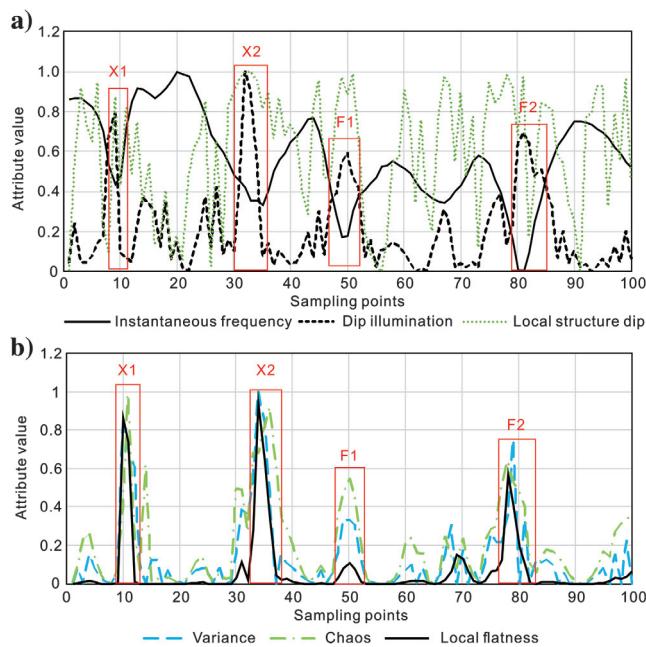
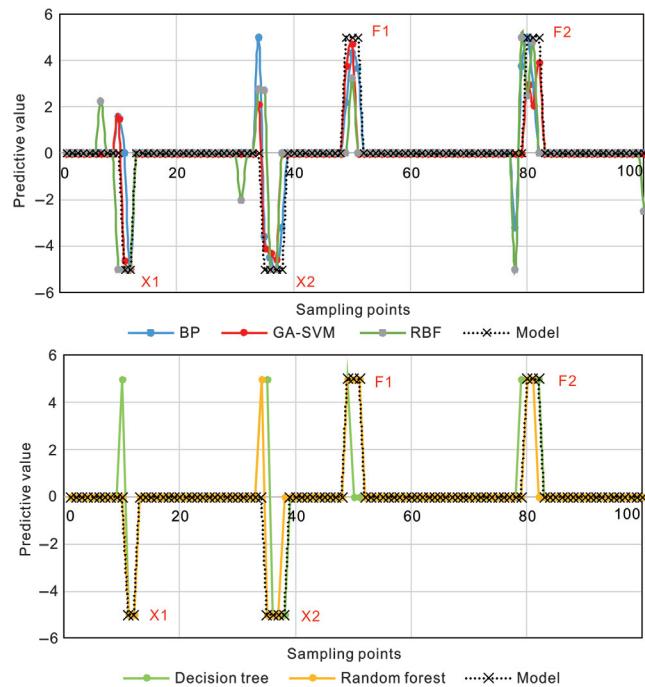
Model-based machine-learning algorithm calculations

Based on the forward data, we extract 100 sampling points along the coal seam, each of which has five attributes, used for the model training of the above five machine-learning algorithms. And, we use the cross-check method to verify the algorithms in which we use 99 sets of data for model training to detect the one remaining data set.

The five network models of the machine-learning algorithm are compared, and the accuracy of the predictions is analyzed (Figure 7). The input layer data of the model are the five sensitive seismic attributes. To achieve classification,

Table 2. Seismic volume attributes (reference from Schlumberger Petrel 2015©).

Attributes	Geologic significance	Definition
3D curvature	Structural continuity	For a particular point on a curve, its curvature is defined as the rate of change of direction of a curve.
Local structural dip	Structural continuity	The estimation of local dip from the seismic data.
Dip illumination	Structural continuity	A new dip estimation method and displays the calculations in two different views.
First derivative	Structural continuity	The first time derivative of the input seismic volume.
Second derivative	Structural continuity	The second time derivative of the input seismic volume.
Variance	Structural continuity, lithology	The estimation of local variance in the signal.
Amplitude contrast	Structural continuity, stratum thickness, lithology, fluid difference	This attribute is loosely based on a Sobel filter, and it attempts to extract and highlight structures in the seismic.
Chaos	Structural continuity, fluid difference	A measure of the lack of organization in the dip and azimuth estimation method.
Cosine of phase	Structural continuity	The cosine of the instantaneous phase, also known as the normalized amplitude.
Dominant frequency	Structural continuity, stratum thickness, lithology, fluid difference	Calculated as the hypotenuse between the instantaneous frequency and the instantaneous bandwidth.
Instantaneous frequency	Structural continuity, stratum thickness, lithology, fluid difference	Calculated from the temporal rate of change of the instantaneous phase.
Instantaneous quality	Structural continuity, stratum thickness, lithology, fluid difference	The instantaneous quality factor is the ratio of instantaneous frequency to twice the instantaneous bandwidth.
Local flatness	Structural continuity, lithology	The variance of the orientation field to identify the uniformity of the signal within the orientation estimation range.

**Figure 6.** The attributes extracted along the coal seam of the forward model.**Figure 7.** Prediction results along the coal seam of the forward model.

we use “−5” for the collapsed column feature and “5” for the fault feature to build the model.

The BP, RBF, and GA-SVM networks are trend predictions. The values change at the position of the structure. So we set the value of the prediction result between [−1, 1] and 0, and we assume that there is no structure. The results of the decision tree and random forest algorithms are only 5, −5, and 0, and no changes are required. It can be seen in Figure 8 that several

algorithms generally predict that a fault exists at the edge of the collapse column, the reason being that the edge of the collapse column of the coal seam is subjected to a downward pulling force, causing the coal seam to be broken or deformed.

A comprehensive comparison of the advantages and disadvantages of each prediction method as well as prediction accuracy and calculating time (Table 3) is helpful to rational selection of an algorithm for different geologic

tasks, work areas, and geologic structure interpretation targets. This would help us achieve efficient and high-precision seismic interpretations. The prediction method with the highest correct rate (97%) is the random forest, whereas the lowest correct rate is 88% for RBF. The fastest calculation time (2.7 s) is the decision tree, whereas the slowest calculation time is 12 s for GA-SVM. The random forest algorithm, when generating each tree, only randomly selects a few features and generally takes \sqrt{m} by default (m means the total number of features). This ensures the randomness of the feature data, so it has better resistance to overfitting. The random forest method shows the advantage of a strong anti-interference ability, and it does not produce overfitting. Its prediction results suggest that this method is the best of the five machine-learning algorithms.

Application

The study area is located in the Shanxi Province, China. The #2 and #5 coal seams are the main coal seams. The basic overview of the study area is shown in Figure 8. It is located in the

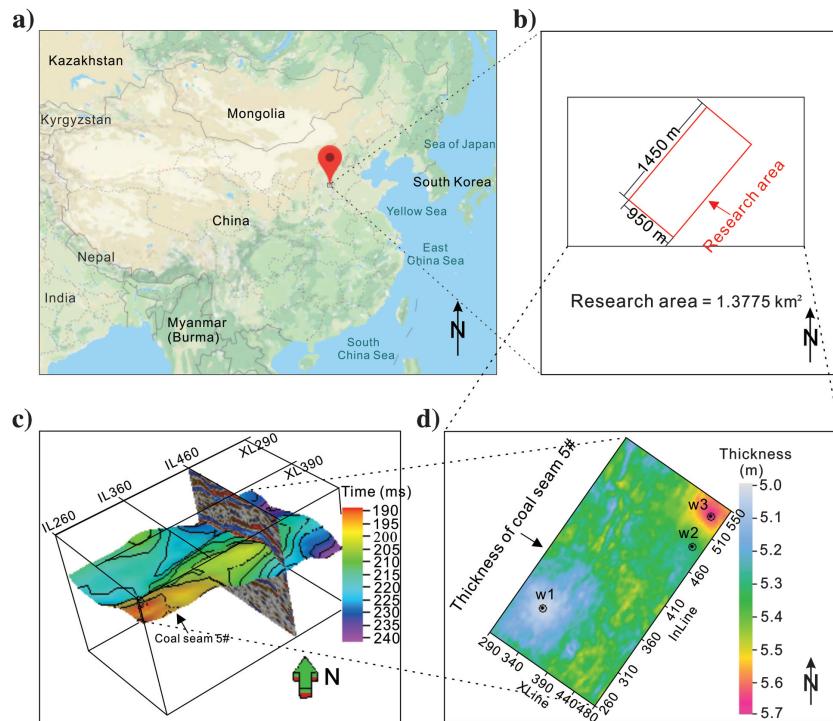


Figure 8. Overview of the study area: (a) geographical location of the study area, (b) study area boundary and size description, (c) contour map of the #5 coal seam time of the target coal seam, and (d) thickness of the #5 coal mine.

Table 3. Comparison of machine-learning algorithms (Mitchell, 2003).

Algorithm	Advantage	Disadvantage	Calculating time (s)	Correct rate (%)
BP	<ul style="list-style-type: none"> High classification accuracy Strong fault tolerance Associative memory function 	<ul style="list-style-type: none"> More parameters required Learning process cannot be observed 	3.4	94
GA-SVM	<ul style="list-style-type: none"> Can solve high-dimensional problems Improves generalization ability 	<ul style="list-style-type: none"> Low computational efficiency Nuclear function parameter is difficult to choose 	12	96
RBF	<ul style="list-style-type: none"> No need to rely on the entire data High computational efficiency Strong global approximation 	<ul style="list-style-type: none"> High sample quality requirements 	2.9	88
Decision tree	<ul style="list-style-type: none"> Simple to calculate and easy to understand Can handle samples of exact attributes 	<ul style="list-style-type: none"> Prone to overfitting Ignores correlation between data 	2.7	94
Random forest	<ul style="list-style-type: none"> Does not produce overfitting Strong anti-interference ability 	<ul style="list-style-type: none"> Similar decision trees may appear Small data do not produce the best classification 	2.8	97

central part of China, as shown in Figure 8a, and it covers an area of approximately 1.3 km² (Figure 8b). The target coal seam is generally high in the south and low in the north, showing a maximum reflection time difference of approximately 50 ms (Figure 8c). The #5 coal seam is the target zone, and the whole area is continuous, with a coal seam thickness of approximately 5 m (Figure 8d). The coal-bearing strata are the Carboniferous Upper Taiyuan Formation and the Permian Lower Shanxi Formation. The faults and subsidence columns in the area are relatively developed, and the scales of development are different. There are large faults in the gap. Two large faults have been identified, with a breaking distance of approximately 20–30 m.

Before coal mining, the roadway will be drilled first, so that actual structure information will be obtained in the study area or adjacent area. Therefore, in the model training, using the actual construction instead of the manual interpretation results will improve the prediction accuracy. At the same time, the first objective of the intelligent interpretation of coalfields is to improve the efficiency of interpretation and the second objective is to improve the accuracy of interpretation. So this paper uses two examples to verify the feasibility of applying machine-learning algorithms to fault and collapse column detection. Example 1 is a 2D planar prediction that uses the actual structures for model training. Example 2 is a 3D prediction that uses a manually interpreted structure for model training.

Example 1

The most accurate random forest algorithm is used to predict the target coal seam (#5) 2D plane. First, the attribute values were extracted along the target coal seam (#5), and five kinds of seismic attributes were obtained (Figure 9). To average the ability of each attribute

value to influence the prediction result, the five seismic attributes are normalized to [0, 1]. Second, the attribute values were selected along the roadway (the black line in Figure 10) of the study area and the exposed geologic structure was assigned. The collapse column was assigned a value of -5, the fault was assigned a value of 5, and the unstructured position was assigned a value of "0." We picked a point every 5 m (one common depth point interval) along the roadway to gather training sample data for the predictive network. A total of 5559 points were sampled in this study area to complete the training of the network model. Finally, all of the attribute point data (55,555 points) over the entire study area were inputted to obtain the prediction result (Figure 10). The total calculation time is 152 s.

In Figure 10, the black line indicates the location of the coal mine roadway and the geologic structure passing

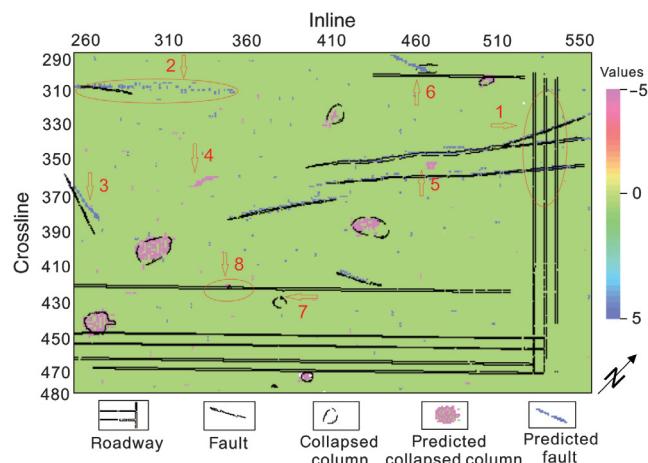


Figure 10. Geologic structure prediction results of the random forest method in the target coal seam.

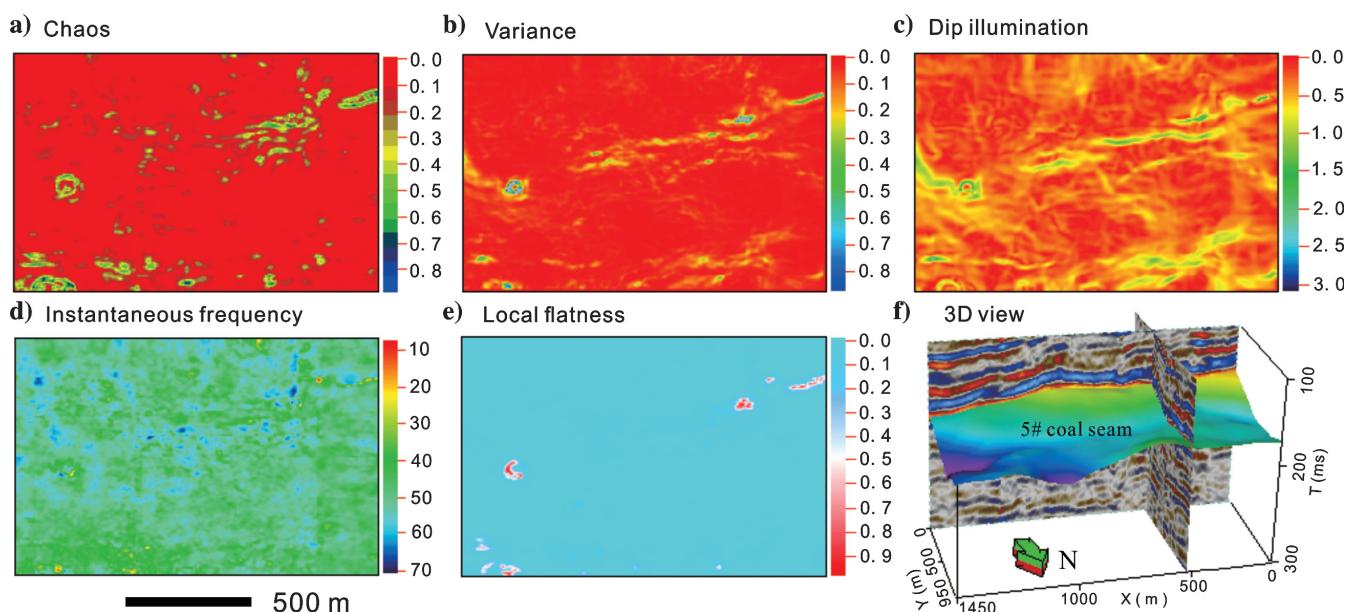


Figure 9. Attributes extracted along the target coal seam (#5).

through the roadway position is the exposed structure. The value of the green part is 0, indicating the absence of a geologic structure, and the value of the dark-green part is 5, indicating the presence of a fault. The value of the gray part is -5, indicating the existence of a collapsed column structure.

The analysis of the predicted results shows that most of the predictions fit the actual observations (Figure 10). The predicted position is in full compliance with the three faults that the roadway passes through (arrow 1). For arrow 2, the conventionally explained fault extension length is short, and the prediction results indicate that the fault should extend. For arrow 3, the predicted fault deviates from the manual interpretation. Note that the results of the random forest method indicate two new collapse column structures and a fault at arrow 4, arrow 5, and arrow 6, respectively. From a geologic point of view,

the development of the collapse column in the mining area is parallel and equidistant along a certain direction, which is highly possible for the collapse column at arrows 4 and 5. For arrow 7, we noted the presence of a collapsed column by manual interpretation, but the prediction by the random forest method indicates that no structures are present. The accuracy of the prediction results for arrows 6 and 7 remains to be verified.

Figure 11 is an enlarged view of a part of the area circled by the red line at arrows 1 and 8 in Figure 10. First, it is notable that for the three faults passing through the roadway shown in Figure 11a, the prediction result is completely consistent with the known positions. Second, in Figure 11b, the prediction result of the random forest method matches with the known collapsed column structure along the roadway. However, it is difficult to interpret the result manually because the size of

the collapse column is very small and the diameter (short axis \times long axis) is approximately 12 \times 15 m.

In Figure 12, several typical structures that pass through the roadway are shown in terms of the arbitrary line in blue, which were interpreted manually. The points in pink and purple depict the random forest method's prediction effects for the collapsed column and faults, respectively. Figure 11 indicates that the predicted fault and collapse column using the random forest method are consistent with the manual interpretation. The positions shown in (1 and 2) are known as collapsed columns, and the positions shown in (3–5) are known faults that the roadway passes over. The prediction results of the random forest method are completely consistent with the manual interpretations. However, positions (2 and 4) mark the small collapsed columns and small faults, respectively. The cross section shows no abnormalities, and the manual interpretation does not show these features. Moreover, for position (3), the manual interpretation deviates from the known structure by approximately 15 m. Therefore, machine-learning algorithms have the ability to recognize smaller structures as well as provide higher accuracy.

Example 2

Based on the random forest algorithm, the same process is used to predict the actual 3D seismic data in the study area. Considering the calculation time, we select the data including the #5 target coal seam in the time direction (150–250 ms). First, because the continuity of other strata except the coal

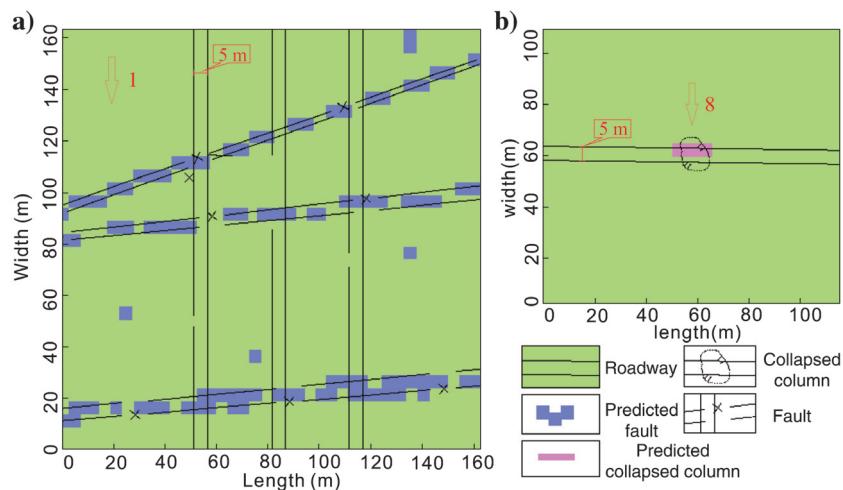


Figure 11. Partial enlargement of the prediction results: (a) magnified view of the structure shown by the red circle at arrow 1 in Figure 10 and (b) magnified view of the structure shown by the red circle at arrow 8 in Figure 10.

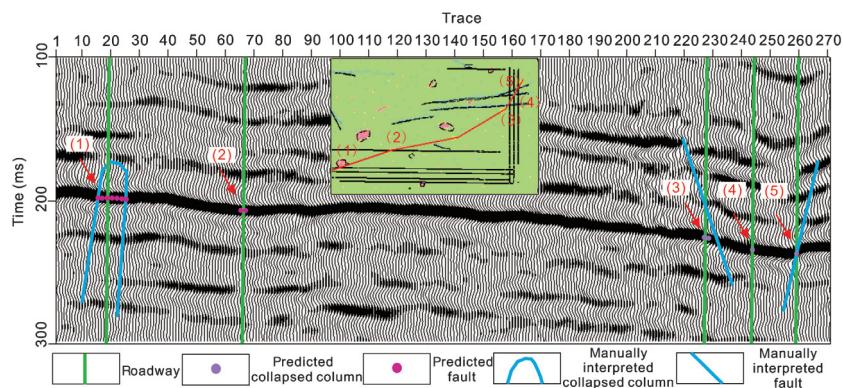


Figure 12. Arbitrary line profile, where the green image in the middle shows the plane position of the arbitrary line (direction from left to right). This prediction is only for the #5 target coal seam. The predicted fault and the predicted collapsed column are represented by the purple and pink points, respectively. The blue line indicates the results of the manual interpretation, and the green line denotes the position of the roadway. Arrows (1 and 2) show the collapsed columns, and arrows (3–5) show the faults.

seam is poor, the original seismic data are smoothed by the local structure to increase the continuity of the seismic reflectors. Second, we extract some of the manual interpretation results as samples of model training (62,377 samples), and we apply this model to predict the actual data (5,611,055 points). The prediction results are shown in Figures 13 and 14. The vertical development of the structures can be clearly seen from the section. As shown, the prediction results are consistent with the results of the manual interpretation; at the same time, some small structural features were also detected. In addition, it also predicts some geologic anomalies that are not caused by geologic structures. From the seismic section, the reflection of the target coal seam is strong and continuous, but the lateral continuity of other strata is poor, and the reflection is disorderly, which is affected by the erosion. Some of the

structural predictions of these strata are not true structures, but they are reflections caused by stratigraphic annihilation. In summary, the machine learning is a valuable tool for seismic interpreters, which can improve the accuracy and efficiency of seismic interpretation.

The two examples use two different workflows to train the model with different sample data. Of course, the more actual structural data, the more favorable it is to create a good model. The workflow of example 1 has three advantages. First, the influence of other layers of disordered reflection can be avoided. Second, the known construction information can be used to improve the accuracy. Third, the calculation time can be reduced. It is suitable for use in the mining stage of a single coal seam. The workflow of example 2 can detect the development of the structures from the 3D space, and it is beneficial to grasp the shape, scale, and extension direction of the

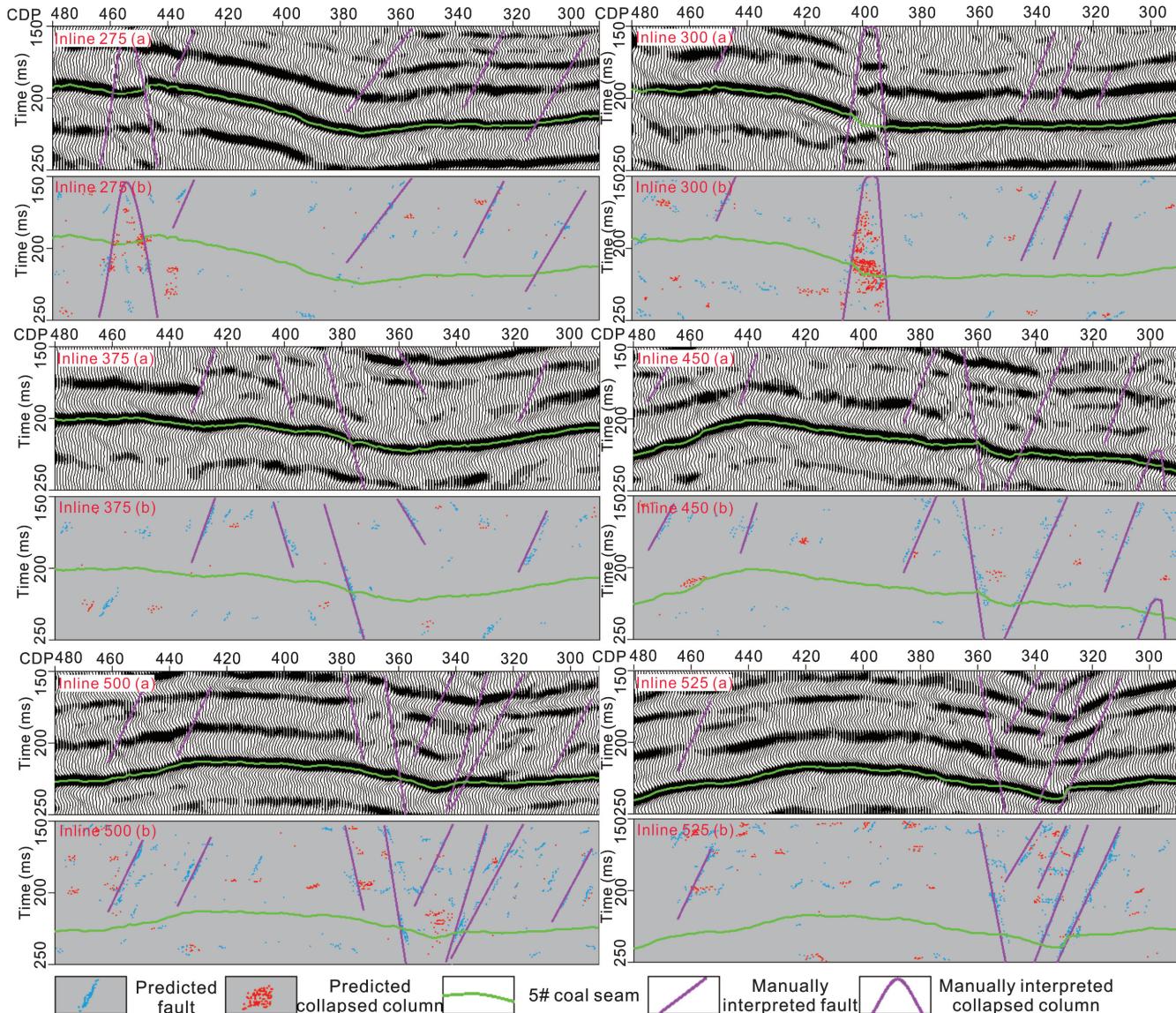


Figure 13. The prediction results at inline 275, 300, 375, 450, 500, and 525. (a) Original seismic section with manual interpretation and (b) prediction results section with manual interpretation.

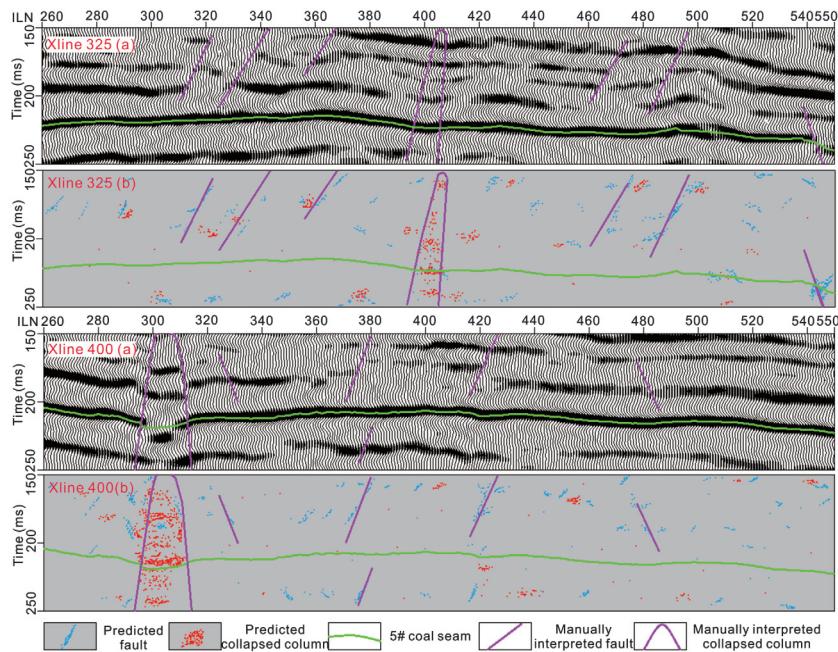


Figure 14. The prediction results at crossline 325, 400. (a) Original seismic section with manual interpretation and (b) prediction results section with manual interpretation.

structures. However, when the manual interpretation is inaccurate, there will be errors in the prediction results, and, second, when the amount of data is large, more calculation time is required.

Suggestions for further study

The seismic structure interpretation method based on machine learning of this paper has high requirements for original seismic data quality and attribute values. So improving machine-learning algorithms and improving algorithm fault tolerance are directions for future research. Seismic attributes not only highlight structural features, but they also amplify noise. Building a good model requires a lot of sample data, which is the disadvantage of machine-learning algorithms. The recently emerging CNN is a deep-learning algorithm that can directly input seismic data. In future research, this algorithm can be tried.

Conclusion

Applying machine-learning algorithms to the prediction of geologic structures can effectively improve the efficiency of seismic interpretation compared to manual interpretation. Forward modeling showed that five kinds of seismic attributes, namely, instantaneous frequency, dip illumination, variance, chaos, and local flatness, are sensitive to faults and collapse columns. For the prediction results of forward model data, the random forest algorithm showed the highest precision (97%) with regard to seismic interpretation. In practical applications, the prediction results are highly consistent with known geologic structures. Thus, the random forest algorithm realized rapid and accurate identification of geologic structures for seismic interpretation.

Acknowledgments

The authors would like to thank the journal editor and reviewers for their constructive suggestions. Thanks are also due to the equipment and financial support provided by the State Key Lab of Coal Resources and Safe Mining. Thanks to the National Key Research and Development Program (grant no. 2016YFC0501102), the National Science and Technology Major Project (grant no. 2016ZX05066-001), and the Coal United Project of National Natural Science Foundation (grant no. U1261203). The authors are grateful to Schlumberger for providing the Petrel software for this research. Warm gratitude is also extended to the members of the research team for their assistance in researching and promoting this work.

Data and materials availability

Data associated with this research are available and can be obtained by contacting the corresponding author.

References

- AlRegib, G., M. Deriche, Z. Long, H. Di, Z. Wang, Y. Alaudah, M. Shafiq, and M. Alfarraj, 2018, Subsurface structure analysis using computational interpretation and learning: A visual signal processing perspective: *IEEE Signal Processing Magazine*, **35**, 82–98, doi: [10.1109/MSP.2017.2785979](https://doi.org/10.1109/MSP.2017.2785979).
- Anees, M., 2013, Seismic attribute analysis for reservoir characterization: 10th Biennial International Conference and Exposition, 119.
- Breiman, L., 2001, Using iterated bagging to debias regressions: *Machine Learning*, **45**, 261–277, doi: [10.1023/A:1017934522171](https://doi.org/10.1023/A:1017934522171).
- Chopra, S., and K. J. Marfurt, 2005, Seismic attributes: A historical perspective: *Geophysics*, **70**, no. 5, 3SO–28SO, doi: [10.1190/1.2098670](https://doi.org/10.1190/1.2098670).
- Chopra, S., and K. J. Marfurt, 2008, Seismic attributes for stratigraphic feature characterization: 78th Annual International Meeting, SEG, Expanded Abstracts, 1590–1594, doi: [10.1190/1.3059386](https://doi.org/10.1190/1.3059386).
- Chopra, S., and K. J. Marfurt, 2009, Detecting stratigraphic features via cross plotting of seismic discontinuity attributes and their volume visualization: *The Leading Edge*, **28**, 1082–1089, doi: [10.1190/1.3236378](https://doi.org/10.1190/1.3236378).
- Chopra, S., and K. J. Marfurt, 2012, Seismic attribute expression of differential compaction: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: [10.1190/segam2012-1323.1](https://doi.org/10.1190/segam2012-1323.1).
- Dezfoolian, M. A., M. A. Riahi, and A. Kadkhodaie-Illkhchi, 2013, Conversion of 3D seismic attributes to reservoir hydraulic flow units using a neural network approach: An example from the Kangan and Dalan carbonate

- reservoirs, the World's largest non-associated gas reservoirs, near the Persian Gulf: *Earth Sciences Research Journal*, **17**, 75–85.
- Di, H., Z. Wang, and G. AlRegib, 2018, Seismic fault detection from post-stack amplitude by convolutional neural networks: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, Tu D 11, doi: [10.3997/2214-4609.201800733](https://doi.org/10.3997/2214-4609.201800733).
- Dong, L., X. Li, and G. Xie, 2014, Nonlinear methodologies for identifying seismic event and nuclear explosion using random forest, support vector machine, and Naive Bayes classification: *Abstract and Applied Analysis*, **12**, 1–8.
- Hibert, C., D. Michéa, F. Provost, J. P. Malet, and M. Geertsema, 2017, Automated seismic detection of landslides at regional scales: A random forest based detection algorithm: *AGU Fall Meeting Abstracts*.
- Hosseini, A., M. Ziaii, A. Kamkar Rouhani, A. Roshandel, R. Gholami, and J. Hanachi, 2011, Artificial intelligence for prediction of porosity from seismic attributes: Case study in the Persian Gulf: *Iranian Journal of Earth Sciences*, **3**, 168–174.
- Huang, L., J. Huang, and W. Wang, 2018, The sustainable development assessment of reservoir resettlement based on a BP neural network: *International Journal of Environmental Research and Public Health*, **15**, 146, doi: [10.3390/ijerph15010146](https://doi.org/10.3390/ijerph15010146).
- Huang, Y., D. Wu, Z. Zhang, H. Chen, and S. Chen, 2017, EMD-based pulsed TIG welding process porosity defect detection and defect diagnosis using GA-SVM: *Journal of Materials Processing Technology*, **239**, 92–102, doi: [10.1016/j.jmatprotec.2016.07.015](https://doi.org/10.1016/j.jmatprotec.2016.07.015).
- Jamaludin, S. N. F., M. Mubin, and A. H. Abdul Latiff, 2017, Imaging of karst on buried Miocene carbonate platform: *Earth and Environmental Science*, **88**, 012011, doi: [10.1088/1755-1315/88/1/012011](https://doi.org/10.1088/1755-1315/88/1/012011).
- Koson, S., P. Chenrai, and M. Choowong, 2014, Seismic attributes and their applications in seismic geomorphology: *Bulletin of Earth Sciences of Thailand*, **6**, 1–9.
- Lee, M. W., T. S. Collett, and T. I. Inks, 2009, Seismic-attribute analysis for gas-hydrate and free-gas prospects on the North Slope of Alaska, in T. Collett, A. Johnson, C. Knapp, and R. Boswell, eds., *Natural gas hydrates — Energy resource potential and associated geologic hazards: AAPG Memoir 89*, 541–554.
- Liu, H., and Y. Jiao, 2011, Application of genetic algorithm-support vector machine (GA-SVM) for damage identification of bridge: *International Journal of Computational Intelligence and Applications*, **10**, 383–397, doi: [10.1142/S1469026811003215](https://doi.org/10.1142/S1469026811003215).
- Liu, J., and K. J. Marfurt, 2007, Instantaneous spectral attributes to detect channels: *SEG*, 23–31.
- Maleki, S., H. R. Ramazi, R. Gholami, and F. Sadeghzadeh, 2015, Application of seismic attributes in structural study and fracture analysis of DQ oil field, Iran: *Egyptian Journal of Petroleum*, **24**, 119–130, doi: [10.1016/j.ejpe.2015.05.001](https://doi.org/10.1016/j.ejpe.2015.05.001).
- Marjanović, M., M. Krautblatter, B. Abolmasov, U. Đurić, C. Sandić, and V. Nikolić, 2018, The rainfall-induced landsliding in Western Serbia: A temporal prediction approach using decision tree technique: *Engineering Geology*, **232**, 147–159, doi: [10.1016/j.enggeo.2017.11.021](https://doi.org/10.1016/j.enggeo.2017.11.021).
- Mitchell, T. M., 2003, *Machine learning*: McGraw-Hill.
- Oyeyemi, K. D., and A. P. Aizebeokhai, 2015, Seismic attributes analysis for reservoir characterization; offshore Niger Delta: *Petroleum and Coal*, **57**, 619–628.
- Rostami, A., M. Arabloo, M. Lee, and A. Bahadori, 2018, Applying SVM framework for modeling of CO₂ solubility in oil during CO₂ flooding: *Fuel*, **214**, 73–87, doi: [10.1016/j.fuel.2017.10.121](https://doi.org/10.1016/j.fuel.2017.10.121).
- Routray, S., A. K. Ray, C. Mishra, and G. Palai, 2018, Efficient hybrid image denoising scheme based on SVM classification: *Optik*, **157**, 503–511, doi: [10.1016/j.ijleo.2017.11.116](https://doi.org/10.1016/j.ijleo.2017.11.116).
- Shamsuddin, A., D. Ghosh, and S. Shahar, 2017, Geological features based on spectral decomposition: Techniques and examples from the Malay Basin: *Earth and Environmental Science*, **88**, 012002, doi: [10.1088/1755-1315/88/1/012002](https://doi.org/10.1088/1755-1315/88/1/012002).
- Waldehand, A. U., J. A. Charles, J. G. Leiv, and S. A. H. Schistad, 2018, Convolutional neural networks for automated seismic interpretation: *The Leading Edge*, **37**, 529–537, doi: [10.1190/tle37070529.1](https://doi.org/10.1190/tle37070529.1).
- Wang, H., T. Wang, Y. Zhou, L. Zhou, and H. Li, 2018a, Information classification algorithm based on decision tree optimization: *Cluster Computing*, 1–10.
- Wang, Z., and G. AlRegib, 2017, Interactive fault extraction in 3-D seismic data using the Hough Transform and tracking vectors: *IEEE Transactions on Computational Imaging*, **3**, 99–109, doi: [10.1109/TCI.2016.2626998](https://doi.org/10.1109/TCI.2016.2626998).
- Wang, Z., H. Di, M. Shafiq, Y. Alaudah, and G. AlRegib, 2018b, Successful leveraging of image processing and machine learning in seismic structural interpretation: A review: *The Leading Edge*, **37**, 451–461, doi: [10.1190/tle37060451.1](https://doi.org/10.1190/tle37060451.1).
- Wang, Z., Y. H. Shao, L. Bai, and N. Y. Deng, 2015, Twin support vector machine for clustering: *IEEE Transactions on Neural Network Learning System*, **26**, 2583–2588, doi: [10.1109/TNNLS.2014.2379930](https://doi.org/10.1109/TNNLS.2014.2379930).
- Wrona, T., I. Pan, R. L. Gawthorpe, and H. Fossen, 2018, Seismic facies analysis using machine-learning: *Geophysics*, **83**, no. 5, O83–O95, doi: [10.1190/geo2017-0595.1](https://doi.org/10.1190/geo2017-0595.1).
- Zhang, C., C. Frogner, M. Araya-Polo, and D. Hohl, 2014, Machine-learning based automated fault detection in seismic traces: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.20141500](https://doi.org/10.3997/2214-4609.20141500).
- Zhao, T., V. Jayaram, K. Marfurt, and H. Zhou, 2014, Lithofacies classification in Barnett shale using proximal support vector machines: 84th Annual International Meeting, SEG, Expanded Abstracts, 1491–1495, doi: [10.1190/segam2014-1210.1](https://doi.org/10.1190/segam2014-1210.1).

Biographies and photographs of the authors are not available.