Automatic Deep Vector Learning Model Applied for Oil-Well-Testing Feature Mining, Purification and Classification

XIN FENG1, QIANG FENG2, SHAOHUI LI3, XINGWEI HOU1, MENGQIU ZHANG1, AND SHUGUI LIU1

1State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China
2Engineering Technology Research Institute, CNPC Bohai Drilling Engineering Company Ltd., Tianjin 300457, China
3National Centre of Metrization for Equipments of Water Transport, Tianjin Research Institute of Water Transport Engineering, Tianjin 300456, China

Corresponding author: Shugui Liu (sgliu@tju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51375338, in part by the Research on Key Measurement Standards and Traceability Technology of Water Transport Engineering, National Key Research and Development Plan, under Grant 2018YFF0212200, and in part by the Special Fund for Basic Scientific Research Business Expenses of Central Public Welfare Scientific Research Institutes under Grant TK202000202.

ABSTRACT Well-testing stage analysis is an important way for oilfield operation state decision-making and reservoir management. However, due to the variability and nonlinearity of the downhole data caused by the complex exploration activities and the differences of petroleum type and geological conditions, the classical methods are ineffective in feature extraction, learning network construction, and classifier optimization. In this work, we propose a new well-testing stage classification method based on a deep vector learning model (DVLM). The novelty of this study lies in the combination of multi-feature extraction, deep learning and feature vector mapping. The proposed method can overcome the problems of poor feature representation ability and poor classification model generalization ability in the existing machine learning methods, which mainly caused by the non-optimized training network structure and the unreasonable classifier design.

Firstly, the initial features are obtained by four classical methods. Then a five layers deep belief network embedded with the mutual information coefficient method is implemented for further feature extraction and purification. Finally, the optimized learning vector quantization classifier outputs the predicted tags.

For model training and testing 572 field samples total of 4004 data streams are used. By considering the classification errors and accuracy metrics, the neurons number of deep learning network and the classifier are tuned, and an optimal and stable framework is obtained. Comparative experimental results with several classical integration models show that the proposed model achieves the highest classification accuracy of 98.065% as well as the least of features (nine). The results demonstrate that the proposed model has excellent performance in improving the classification accuracy and completing the feature compression. Moreover, the proposed model has very important practical significance for guiding the automatic analysis and processing oil and gas data.

INDEX TERMS Well-testing stage analysis, feature extraction, deep belief network, learning vector quantization.

I. INTRODUCTION

During oil well exploitation, the accurate classification of the well-testing stage plays an important role in real-time early warning of the operating platform, the rational revision of the production flow and the scientific management of the reservoirs [1]. The off-line well-testing stage classification can improve the well-testing operation process, provide qualitative explanations to the data without assignment of the exploration category and working stage, and complete the accurate evaluation of complex oil reservoirs. The online well-testing stage classification can determine whether an operation stage is finished or terminated in advance, reduce production accidents, and provide guidance for real-time...
surface operations. However, the high nonlinearity of geological parameters and the high complexity of exploration processes caused by the continuous development of oil exploration in depth and width make it difficult for the existing classical mathematical method to extract a large amount of effective information in the exploration data.

Moreover, the testing data contains a lot of interference signals, singular signals, and information about the ground operations and the activities of the staff due to the temporary operation flow adjustment or the occurrence of uncertain events at strata and borehole locations [2]. Even when the data is from the same operation stage but different testing wells, there exists a large difference in the length and the curve shape. These uncertainties exacerbate the difficulty of analyzing and automatically processing the massive data [3], which pose significant challenges to data analysis and directly lead to the inaccurate matching relationship between testing data and operation stage. Therefore, the problem of “rich sample data but poor effective information” often occurs.

At present, the unsatisfactory data classification results are mainly attributed to poor characterizing information ability of data features and low generalization ability of classification models. In order to solve this problem, it is urgent to mine a few but high representational characteristics of the well-testing data that would facilitate the effective classification of non-uniqueness pressure signals.

As for improving the characterizing information ability of data features, the time-domain approach has long been popular because of its simplicity and ease of implementation. Numerous methods have been proposed for the feature extraction and classification in the field of petroleum operation decision-making and system diagnosis [4], such as coastal oil tank detection [5], oil spill detection [6], pipeline leakage detection [7], tool wear prediction [8], and working status inspection [9]. Li and Misra [10] used nuclear magnetic resonance to investigate the internal structure of geomaterials filled with fluid. Firoozabadi et al. [11] built a well-prediction model based on a pressure drop. Meckel et al. [12] conducted the wellbore leakage model optimization by analyzing reservoir pressure transients. Ghaffarian et al. [13] illustrated that transient pressure data could be used to identify the correct reservoir models with high accuracy. It should be noted that the pressure derivative has always been an important time-domain analysis method used in the modern well test analysis to obtain the properties and the conditions of the flow system [14]. Miao and Di [15] took pressure derivative as feature to track and evaluate five kinds of fracturing methods commonly used in oil production plants. However, the pressure derivative represents the rate of pressure change, and can only reflect the local features rather than the global features. Besides, some studies also tried the nonlinear methods. Ahmadi et al. [16] proposed a well-testing identification model for the pressure transient test data using the concept of shapelets. Li and Misra [10] extracted the features by a three-level wavelet packet transform (WPT) from the resultant force converted from thrust and torque. Zheng et al. [17] analyzed the real-time mud signals and designed a feature extraction method based on the intrinsic mode functions through the ensemble empirical mode decomposition (EMD). Yang and Kong [18] used the combination method of the wavelet transform and EMD to detect mutation point of the positive pressure wave signal. Although the above-mentioned studies have achieved a certain interpretation effect, their feature extraction algorithms are single and shallow and the simple single feature they extracted cannot characterize the trend of the data to be processed. We are pleased to find that the multi-feature extraction method could resolve the above problems and provide guidance for the follow-up study [19].

We notice that the intelligent technologies have been widely used in many fields. As one of these methods, deep learning can not only extract the effective features of data, but also can be used for prediction and classification [20], it shows good performance in text recognition [21], manufacture diagnosis [22], fault detection [23], electrocardiogram (ECG) detection [24], and speech recognition[25]. Hammad et al. [26] employed a pre-trained deep CNN models and selected valuable layers to get a good representation of ECG and fingerprint data. Kim et al. [27] created a particle swarm optimization–deep belief network (PSO-DBN) based classifier, which is used for rare class prediction. Zou et al. [28] constructed a feature-selection model to improve the remote sensing scene classification performance, in which DBN was used for mass-data processing and feature reconstruction. The most recent work using deep learning in well-testing applications has been presented by Feng et al. [19], in which he created an integrating model based on deep belief network for the feature extraction. The model is almost the first to introduce the Boltzmann machine based deep learning methodology into well-testing data analysis. Although the paper provides the optimal structure of the DBN network suitable for the division of the oil testing data, the step size of neurons in DBN highest layer was 10, which was a bit large. At the same time, the support vector machine (SVM) classifier used in his model needs to be normalized, which requires higher format requirements for input features.

The above-mentioned studies illustrate the advantages of the DBN method in processing the data and enhancing the classification ability of models under different application backgrounds, but these methods still have many drawbacks. First, they lack the consideration of the impact of the DBN structure on data analysis, a complex learning network may cause too many training sessions, resulting in low learning efficiency and slow convergence. Second, there is no general theoretical guidance on how to determine the number of hidden layers and the number of nodes in the hidden layer of the learning network, and it mainly draws on a lot of practical experience. Third, Restricted Boltzmann Machines (RBM) and the back propagation (BP) network, as the important components of DBN, still have the disadvantages that they are easy to fall into local optimality and cannot obtain
global optimal solutions. Fourth, the feature redundancy still exists. Hence, further improvement in features purification and model optimizing are required.

As for improving the generalization ability of classification models, scholars have been studying this problem by using classic methods such as Naive Bayesian (NB) [29], RUSBoost [30], probabilistic neural network (PNN) [31], k-nearest-neighbor (KNN) [32], but so far there is no universal solution. In the latest method, Shahriari et al. [33] tried to illustrate the performance of the deep neural network (DNN) for inverting logging-while-drilling measurements acquired in high-angle wells. In our previous research [19], the SVM with PSO optimized parameters was only used for data classification. The learning vector quantization (LVQ) [34] is a classic vector quantization framework [35] and has both feature learning and classification functions. It has been successfully applied to many classification fields, such as financial management [36], industrial production [37], system safety [38], and wearable physiological parameter detection [39]. It can be used to build the codebook from various code vectors and has the wider input features range that makes it possible to improve the model generalization ability. Furthermore, an LVQ network structure suitable for well-testing analysis is rarely used. Therefore, it is of great practical significance to introduce the LVQ into oil and gas exploration in addition to the classical mathematical and the intelligent methods. The only drawback in LVQ is that the optimal parameters need to be set according to the analysis of the processed data.

The main tasks of this article are extracting the effective features of well-testing data and optimizing the structure of the integration model. When classifying the data obtained under complicated working conditions, the classification accuracy tends to be poor by using single feature extraction or above-described classification methods [40]. Therefore, in order to improve the stage classification performance (that is, the characterizing information ability of data features and the generalization ability of classification models), we propose the deep vector learning model (DVLM). Firstly, four classic feature extraction methods are used to extract the initial multiple features. Next, the DBN is employed to learn, excavate, and reconstruct the features in-depth. Then, the maximum information coefficient (MIC) [41] is applied to delete the data redundancy. Finally, the LVQ is used to classify the features. The main contributions of this article are summarized as follows:

1. We are the first to integrate multi-feature extraction, deep learning and feature vector mapping and use their integration for well-testing working stage classification.
2. We analyze the priority of the number of neurons in the highest level of DBN from 20 to 60 with a step size of 1. Although it belongs to the second study, it supplements the priority rule missed in our previous study due to too large step size (step size = 10).
3. We are the first to apply the cross-validation method based classification error and accuracy on well-testing data to optimize the structure and parameters of the LVQ classifier.
4. We are the first to use MIC method for feature purification and redundant elimination based on LVQ classification rate.

II. WELL-TESTING DATA ANALYSIS

Well testing is a seepage mechanics based method to acquire downhole geological information and production capacity by analyzing the dynamic changes of pressure with time in the production process of oil wells and gas wells. As the theoretical test curve is shown in Fig. 1(a), the complete well-testing operation process consists of total five stages, and the operation curve of each stage is a data-stream, which is composed of the geological parameter pressure of the corresponding time period. In this article, we do not directly classify the geological parameters, but classify the data streams composed of the geological parameters. In order to keep the originality of the well-testing data, the preprocessing of the data streams is only normalization. However, the signal processing circuit of the electronic pressure gauge has rectified and filtered the signal of the sensor when acquiring the pressure data.

Since well opening and well closing are the two most important stages to be identified, in most field operations, these two stages are performed at least twice. After merging the same stages, the whole process consists of a total of five stages. Set the tag value for each stage and use these values for the standard judgment. The surface staff will identify the stage name based on the predicted tag. The five working stages and the detailed correspondence stage tags are shown in Table 1.

| Stage Order | Working Stage       | Stage Name | Stage Tag |
|-------------|---------------------|------------|-----------|
| AB          | lower the oil string| S1         | 1         |
| BC          | waiting stage       | S2         | 2         |
| CE          | first well opening  | S3         | 3         |
| EF          | first well closing  | S4         | 4         |
| FH          | second well opening | S3         | 3         |
| III         | second well closing | S4         | 4         |
| JK          | pull-up the oil string | S5       | 5         |

The conventional well-testing works such as curve analysis and operation stage classification are mostly done manually by experts with production experience and contain a lot of subjectivity. Take the data in lowering the oil string stage as example. As shown in Fig. 1(b), if denosing process is required, the pressure noise points that deviate significantly from the “average” trajectory of other testing points will be quickly identified and filtered artificially due to their obvious fluctuation characteristics. However, the manual methods will not be able to deal with the cases where the point
The extraction of strong characterizing information from data is difficult due to the stochasticity of data and differences in algorithms. During oil and gas exploitation, the procedure of well testing is scheduled, planned, and repeatable, ensuring that the operation stage sequences are not random. However, the interference of outside and artificial factors on well-testing data is inevitable. As shown in Fig. 1(b), the pressure and time are not simple linear relationships. The environmental factors that affect the relationship between the two include the geological conditions of the downhole operation area, the surface operation environment, and the measured downhole depth. These factors lead to the different underground energy and pressure recovery capacity of different test samples, the uncertain running time of each stage, which makes the pressure curve of each operation stage have a great difference in amplitude, shape, and time width.

Besides, during the well-testing signal acquisition process, various interference signals are inevitably produced due to the influence of the mechanical vibration, the faulty or delayed operation and the negligence use of oilfield workers. These noises and uncertain interferences are mixed with the real downhole geological pressure signals, which will lead to distortion of pressure signals and ambiguity of pressure waveforms, thus leading to inaccurate identification of signal feature points and unobvious distinguishable features between the data in each stage. These noises show a rapid rise and fall in the time domain.

As shown in Fig. 1(b), the position, the duration and the occurrence of the peak and the inflection point on the timeline vary significantly in different well-testing sample curves. The density of peak group can be presented in frequency domain. There exists some similar but difficult to extract features from different samples or stages, such as continuous peaks, stepped pressure rise/drop mixing various noises and interferences. Among them, the continuous peaks are mostly generated by screen plugging, packer loss of seal and pressure-relieving operation during the well closing, and the mechanical vibrations. The stepped rise often occurs when the equipment fails, the sampling frequency is too low, and lowering the oil string. The stepped drop is easy to form in variable-rate drawdown test and pulling up the oil string stage. The step width and curve rise/fall slop well reflect the time-frequency characteristics of these stepped rise/falls.

The above similarities and uncertainties show that, due to the influence of the workflow and the geological environment, the data features are difficult to calculate by using simple empirical formulas or mathematical statistical methods. However, the good news is that the time domain methods, frequency domain methods, and time-frequency domain methods can all be used to abstractly characterize the data.

However, the above difficulty in extracting the data features cannot be solved by simple time-frequency feature extraction. On the one hand, the existing classical feature extraction methods cannot cover all the feature types. On the other hand, extracting only one specific feature type using a single method is not enough to fully characterize the information of data. Therefore, to distinguish the changes in production operations such as noise interference, formation blockage, it is necessary to conduct an overall and regional
treatment of the well-testing data. The multi-feature integration method should be utilized to represent all possible types of well-testing features. Meanwhile, there are the following new problems that need to be further solved:

1. Whether the initial features obtained by the multi-feature extraction method are the features that can best characterize the data stream.
2. Whether there is a need to improve the feature representation ability by re-learning the initial features.
3. More features also mean more redundancy, whether they need to be purified.
4. Whether there exists a mapping method that requires little input vector formatting requirement and easily converges various complex and abstract input features to the categories they belong to.

Fortunately, we find all solutions and believe it is necessary to use time or/and frequency domain methods and the deep learning method to characterize the inherent laws of the reservoir parameters, and use the LVQ to classify the heterogeneous characteristics of the well-testing data.

![FIGURE 2. DVLM model framework.](image)

III. METHODOLOGY

To overcome the challenges brought by the above problems, we propose a deep vector learning model, whose framework is shown in Fig. 2. We focus on using the classical methods to classify the complex and nonlinear downhole geological parameters, which are mainly reflected in two parts: mining the data characteristics and improving the generalization ability of the classifier.

In this section, we follow the process of model formulation to explain the working principles of different methods used in the DVLM model. The data processing steps can be concluded as follows:

Step 1: Perform the multi-feature extraction (MFE) by using four classical time-frequency methods and obtain five types of initial features.

Step 2: Reconstruct and purify the initial features by using the deep belief network-maximum information coefficient integration methods.

Step 3: Apply the LVQ classification accuracy to the analysis of the effect of various influence factors on model structure optimization and feature classification, so as to ensure the optimal performance of the model.

A. THE MFE METHOD

Practice confirms that due to the nonlinear-time-varying characteristic of downhole signal, when applying the pure time-domain or frequency-domain features to data classification, it usually encounters the problem of poor feature characterization ability. So learned from MFE method [17], we integrate wavelet packet decomposition-approximate entropy (WPD-AE) [42], empirical mode decomposition-approximate entropy (EMD-AE) [43], fast Fourier transformation (FFT), and linear regression (LR) to process and distinguished the similarities and uncertainties between the different well-testing data.

Define the \( n_{00}^{(l)} \) group data \( s(n_0, N, x, y) \) with data length \( N \), \( y \) is the amplitude of the data, and the index number \( x \in [0, N - 1] \). The energy of the well-testing signal is mostly distributed in the low frequency band. Therefore, we use WPD to perform three-layer decomposition of the well-testing data to obtain the reconstruction coefficients \( \{d_{il}^{(l)}(n_0)\} \), where \( l \in [1, 2^{L-1}] \), and \( C_L \) is the WPD decomposed level. Considering that Symlets has good local characteristics in the frequency domain, and relatively high accuracy in the signal phase reconstructed, we select Symlets for WPD wavelet bases.

EMD is suitable for non-linear and non-stationary data analysis which contains multiple IMF components \( \{c_{nl}^{(l)}(n_0)\} \) and a residual \( r(n_0) \), \( n_l \in Z \). To make the eigenmode function have good linearity and stability, CS is set to 0.2 in the EMD. To avoid the abnormal processing caused by too large data or unequal data length between two data samples, the data length of each sample is sampled to 1000.

Further, we use AE [33] to indirectly describe the complex changes inside the \( \{d_{il}^{(l)}(n_0)\} \) and \( \{c_{nl}^{(l)}(n_0)\} \). Thus, the WPD-AE \( \alpha(n_0) = ApEn_{nl, m}(n_0, d_{il}) \) and the EMD-AE \( \beta(n_0) = ApEn_{nl, m}(n_0, c_{nl}) \) are obtained. For a specified length data sequence, the magnitude of the AE depends on the dimension \( m \) and the tolerance threshold \( r \). To make AE less dependent on \( N \), set \( m = 2 \) and \( r = 0.2SD \), where SD denotes the standard deviation of \( s(n_0, N) \).

FFT is used to extract the peak distribution and sampling frequency of the data. The characteristics \( X(k_0, m_0) \) are obtained by calculating the top \( k_0 \) FFT coefficients.

Furthermore, an oil well interpretation system typically uses pressure derivatives for reservoir evaluation and geological research, but seldom for well-testing stages classification. Faced with the stage data that requires global data analysis to complete classification, the pressure derivative belongs to the local data transient analysis method and is not functional. Production experience shows that gradient features can quickly and accurately predict the motion trend of the detected data. It’s working principle is similar to pressure derivatives. Thus, we use the LR to extract the parameters including three-interval regression (TIR) \( \text{reg}_{3}(n_0) = [r_{a1, b1, c1}, r_{a2, b2, c2}, r_{a3, b3, c3}] \) and multi-interval extreme (MIE)\( \text{Reg}_{nd}(n_0) = [\text{min}(r_{a1, b1, c1}), \text{max}(r_{a1, b1, c1})] \). \( r_{a1, b1, c1} \) (expressed in Eq. (1)) is the gradient value of the \( id^{th} \) interval data by dividing \( s(n_0, N) \) into \( nd \) parts.

\[
\begin{align*}
rd_{nd, il}^{(n_0)} &= \sum_{kd=0}^{(N/nd)-1} (x_i - \bar{x})(y_i - \bar{y}) / \sum_{kd=0}^{(N/nd)-1} (x_i - \bar{x})^2 \\
\bar{x} &= \sum_{kd=0}^{(N/nd)-1} x_{id} \cdot (nd / N) \\
\bar{y} &= \sum_{kd=0}^{(N/nd)-1} y_{id} \cdot (nd / N) 
\end{align*}
\]
where \( nd > 3 \), \( id \in [1, nd] \). \( \bar{x} \) and \( \bar{y} \) represent the mean of index number and the mean magnitude of the \( id \)th interval data, respectively; \( kd = 0, 1, \ldots, (N/nd) - 1 \).

Then, the final multi-feature vector can be expressed as:

\[
\text{attr}(n_0) = [\alpha(n_0), \beta(n_0), X(k_0, n_0), \text{reg}_3(n_0), \text{Reg}_{nd}(n_0)]
\] (2)

Set \( k_0 = 20 \), \( nd = 10 \) and 100. As shown in Fig. 3, the MFE method extract a vector containing 38 feature elements from the \( n_0 \)th sample. For one sample, the process of acquiring MFE features is shown in the appendix at the end of the paper (see Fig. 13).

The output features can be expressed as:

\[
y_{h-\text{level}} = f(\psi; \theta_{\text{DBN}}).
\] (3)

where \( \psi \) is the input feature, \( y_{h-\text{level}} \) is the output feature.

Considering that feature redundancy still exists in \( y_{h-\text{level}} \), we prioritize and purify the \( y_{h-\text{level}} \) based on MIC method. The main steps were as follows.

Step 1: Calculate the MIC value \( M(y_{h-\text{level}}, I_{h-\text{level}}) \) using Minepy package, \( I_{h-\text{level}} \) is the element index in \( y_{h-\text{level}} \).

Step 2: Update the index sorting based on MIC value from large to small and get the priority ranking.

Step 3: Calculate the LVQ classification accuracy of the above rank, and denote the top \( \tau \) features that can obtain the highest accuracy as \( y_{\text{MIC}}(n_0) \) and name it as purified feature.

### C. CROSS-VALIDATION-OPTIMIZED LVQ CLASSIFIER

We use the LVQ1 to design the classifier. The LVQ has no normalization requirements for input features and needs only a few parameters to be optimized. As shown in Fig. 4, the LVQ network is composed of an input layer (IL), a competition layer (CL), and a linear output layer (LOL).

By calculating the CL neuron closest to the input vector, the LOL neuron connected with the CL neuron is found. Define the weight between the \( j \)th neuron of IL and the \( i \)th neuron of CL as \( w_{ij} \), the number of neurons in the CL as \( N_C \). The steps of the LVQ1 can be concluded as follows.

Step 1: Initialize the \( w_{ij} \) and learning rates \( \eta (\eta > 0) \) between IL and CL.

Step 2: Send the input vector \( X = (x_1, x_2, \ldots, x_R)^T \) to IL and calculate the distance \( d_i \) between the neurons of CL and \( x \), where

\[
d_i = \sqrt{\sum_{j=1}^{R} (x_j - w_{ij})^2}, \quad i = 1, 2, \ldots, N_C
\] (4)

Step 3: Define the neuron of CL with the smallest \( d_i \) from \( X \) as \( \text{Min}_c \), mark the class tag of the LOL neuron connected to \( \text{Min}_c \) as \( C_i \).
Step 4: Define the class tag of \( X \) as \( C_X \), \( w_{ij} \) are updated as follows:

\[
\begin{align*}
    w_{ij,\text{new}} &= w_{ij,\text{old}} + \eta(x - w_{ij,\text{old}}), \quad C_i = C_X \\
    w_{ij,\text{new}} &= w_{ij,\text{old}} - \eta(x - w_{ij,\text{old}}), \quad C_i \neq C_X
\end{align*}
\]  

Besides, the cross validation (CV) method [47] plays an important role in avoiding over-fitting and improving the generalization ability [48]. To optimize the LVQ structure, we used 10-fold CV to partition datasets into complementary subsets, ensuring that each example has the same chance of appearing in the training and testing set.

E. PROPOSED MODEL

This article aims at automatically and accurately outputting the tags belonging to the data of unknown classification stage relying on training the oilfield datasets with class tags. Generally, combining the classifications of an ensemble is often more accurate than the individual methods making them up. Therefore, we established the topology structure of the stage classification model based on MFE and DBN, and put forward the DVLM, as shown in Fig.6. The detail works we have done are:

1. Combine the classical feature extraction algorithms and deep learning network to a deep learning system.

We advocated extracting as much feature information as possible through the analysis of well-testing data, which contained the essential information representing all the operation process.

2. The optimization of DBN network structure.

Aiming at the problems of poor learning ability of classical three-layer network and difficulty in determining the number of neurons in the middle layer, we created a self-configuring DBN network and determined the optimal network structure suitable for well-testing data processing.

3. The optimization of classifier parameters.

The CV algorithm was introduced to avoid the local minimum value while seeking the optimal number of neurons in CL of the LVQ classifier in order to obtain global optimal parameters.

4. Large collection of well-testing samples.

 Acquisition of samples require a long operation time due to the fluidity of oil and the slowness of pressure recovery, which brings great challenges to models that rely on massive learning datasets to complete data classification. We collected the samples from single well to well groups, and used the obtained data of past 10 years to enhance the learning and the classification capabilities of the model.

The DBN network in the proposed model is a multi-layer forward network with BP feedback mechanism, and has a five-layer network structure, including one input layer, three RBM hidden layers and one output layer. The layers are fully connected, and there is no connection between the neurons in each layer. The three hidden layers are used for feature learning, reconstruction, and dimension reduction. The hidden layer nodes are used to extract and store the inherent laws of the data. To make the DBN have a strong learning ability, we set the number of input layer neurons to be nearly 10 times the number of input features. Moreover, the number of hidden neurons should be less than twice the size of the input layer [50]. Based on this, we set that the numbers of neurons in the three hidden layers \( R_{\text{hid}} \) decrease layer by layer, the ratio is 2, so the number of neurons in the three layers are 400, 200 and 100 respectively. The previous analysis results showed that the MFE method obtained a vector with 38 multiple features for each sample. Therefore, in Fig. 6, the number of neurons in the first visible layer \( R_{\text{first}} \) was set to 38. The number of neurons in the highest layer \( R_{\text{high}} \) and the LVQ competition layer \( R_{\text{CL}} \) were optimized by analyzing the LVQ classification accuracy under different \( R_{\text{high}} \) and \( R_{\text{CL}} \) (shown in simulation results). Here, the features of the training and the testing sets are expressed as \( \text{Attr}_{\text{train}} \) (2402 × 38) and \( \text{Attr}_{\text{test}} \) (1602 × 38), respectively. For the \( n_{\text{th}} \) sample, the main definitions are listed in the appendix at the end of the paper (see Table 4).

The concrete process and the pseudocode of the DVLM are described in Algorithm 1.

IV. EXPERIMENTAL RESULTS

The proposed model is obtained on the basis of learning and training the sample set with given stage tags, and
FIGURE 5. The well-testing workflow and its essential equipment.

FIGURE 6. The topology structure of DVLM. First, use MFE to process the data sets and obtain a feature vector containing 38 elements. Second, input 38 feature elements into the DBN network to complete the layer-by-layer learning. Third, do MIC purification, the number of purified features is reduced to 9. Fourth, input the purified features to the LVQ classifier. The numbers in parentheses represent the number of neurons contained in the corresponding layer.

The data classification experiments are conducted by using MATLAB 2017a. The classification accuracy $Acc$ is obtained by comparing the predicted stage tags with the actual stage tags of corresponding samples. The expression for accuracy is derived in Eq. (6).

$$Acc = \frac{\sum_{i=1}^{n_{\text{test}}} i_{\text{pred}}}{n_{\text{test}}}, \quad \text{if} \begin{cases} \text{Tag}'(n_0) = \text{Tag}(n_0), & i_{\text{pred}} = 1 \\ \text{Tag}'(n_0) \neq \text{Tag}(n_0), & i_{\text{pred}} = 0 \end{cases}$$ (6)

All the works that need to verify the classification ability are accomplished by the $R_{CL}$-optimized LVQ. Define the cumulative error $e = \sum_{i=1}^{n_{\text{test}}} i_{\text{pred}}$, the goal of LVQ training parameters as $G_{\text{Param}}$. The most challenging of LVQ is its difficulty in selecting an appropriate $R_{CL}$ and $G_{\text{Param}}$. Considering that CV is available for estimating an appropriate parameter value, we randomly selected 60% of samples for training and the remaining for testing to obtain the optimal LVQ parameters. Given the range of $R_{CL}$ values and configuration parameters (epoch and $G_{\text{Param}}$), we first obtain the local optimum number of neurons using 10-CV. In each 10-CV estimation, 10 errors are obtained. After trial and error, the ranges of the global optimal parameters are determined and $R_{CL} \in [10, 50]$, $G_{\text{Param}} \in [0.24, 0.33]$. Let $Err$ in Eq. (7) be the judgment criteria of parameters optimization, where $err_1$ reflects the dispersion, $err_2$ reflects the average distribution. Then the combination of $G_{\text{Param}}$ and $R_{CL}$ with the smallest $Err$ is selected as the optimal combination of parameters.

$$Err = err_1 \cdot err_2 = \sqrt{\sum_{i=1}^{10} (e_i - \bar{e})^2 / 10} \cdot 2 \cdot \sqrt{\sum_{i=1}^{10} (e_i)^2 / 10} / (e_{i, \text{max}} - e_{i, \text{min}})$$ (7)

Fig. 7 shows the distribution of $Err$ at different values of $G_{\text{Param}}$ and $R_{CL}$. When $G_{\text{Param}} = 0.28$, the processed results of LVQ not only have the lowest average error, $Err_{\text{ave}} = 0.2876$, but also contain the minimum $Err (Err_{\text{min}} = 0.2272)$.
Algorithm 1 DVLM Model

**Input:** $s(n_{train})$, $Tag(n_{train})$, $s(n_{test})$.
**Output:** $Tag(n_{test})$.

**Model optimization and $s(n_{train})$ training:**
1. Set $C_L$, $nl$, $nd$, $k_{train}$.
2. **Foreach** $n_{train} \in \{1, \ldots, 2402\}$ **Do**
   3. Get $\alpha(n_{train})$, $\beta(n_{train})$, $X(k_{train}, n_{train})$, $reg_3(n_{train})$, and $Reg_{nd}(n_{train})$.
   4. Get $attr(n_{train}) = [\alpha, \beta, X, reg_3, Reg_{nd}]$.
5. **End Foreach**
6. Get $Attr_{train}(2402 \times 38) = \{attr(n_{train})\}$.
7. Get the optimal $R_{CL}$ based on 10-CV method.
8. Perform the RBM training.
9. Get $\theta_{DBN}$ using the BP network.
10. Get the optimal $R_{high}$ based on LVQ classification results.
11. Calculate $y_{h-level}$ by Eq. 3.
12. Sort $y_{h-level}$ by using the MIC and get $y_{MIC}$.

**Classification of $s(n_{test})$:**
13. **Foreach** $n_{test} \in \{1, \ldots, 1602\}$ **Do**
   14. Perform the steps 3-4 and get $attr(n_{test})$.
   15. Perform the step 6, 8, and 11-12 using $\theta_{DBN}$ and optimal $R_{high}$.
   16. Get $Tag(n_{test})$ based on $R_{CL}$-optimized LVQ classifier.
17. **End Foreach**

For all parameters combinations. So after successive training, $G_{Param} = 0.28$ and $R_{CL} = 30$ are the optimal choice.

We capture the least MSE and the highest LVQ accuracy for the best DBN configuration. From 20 to 60, with a step size of 10, five trial designs are obtained. Fig. 8 shows the training set classification accuracy of the DBN with different numbers of highest-layer neurons. When $R_{high} = 30$, the lowest MSE is 0.0053289, while its LVQ accuracy is as high as 0.972534. Thus, the optimal number of highest-layer neurons that would yield satisfactory performance is determined.

At present, there is no scientific and universal method for determining the structure of deep learning networks. Fig. 9 analyzes the impact of different $R_{high}$ on classification accuracy and gives its priority ranking, where the abscissa represents the number of neurons in the highest layer of DBN. When the value of $R_{high}$ is greater than 41, the accuracy is stable in the range of 97.07%—97.24%, which could ensure sufficiently high network performance. Considering that the too many neuron nodes will increase the calculation burden and appear “overfitting” during parameter optimization [51], the basic principle of choosing optimal $R_{high}$ in this article is: the network structure should be as compact as possible by reducing the number of hidden layer nodes with a high classification accuracy ensured. So, in summary, $R_{high} = 30$ is chosen.

Table 2 shows the classification accuracy before and after the MIC processing at different numbers of highest layer neurons. The self-validation accuracy and MIC classification accuracy are obtained by using $y_{h-level}$ and $y_{MIC}$, respectively. The difference is that both training samples and
TABLE 2. Classification accuracy before and after the MIC processing. DBN∗ represents the DBN training network with * highest-layer neurons.

| Model type                  | DBN20 | DBN30 | DBN40 | DBN50 | DBN60 |
|-----------------------------|-------|-------|-------|-------|-------|
| Number of top-level nodes   | 20    | 30    | 40    | 50    | 60    |
| Self-verification accuracy  | 0.948793 | 0.982931 | 0.954205 | 0.961699 | 0.977102 |
| The MSE of yh−level         | 0.020483 | 0.006828 | 0.018318 | 0.015321 | 0.009159 |
| Number of MIC features      | 10    | 9     | 9     | 33    | 20    |
| MIC classification accuracy | 0.939451 | 0.980649 | 0.945069 | 0.958801 | 0.969413 |
| The MSE of yMIC             | 0.021649 | 0.007660 | 0.016653 | 0.017652 | 0.012490 |
| MIC self-verification accuracy | 0.939158 | 0.983760 | 0.972839 | 0.965678 | 0.982839 |

Verification samples in self-validation were from $y_{h-level}$, while in MIC classification, 60% of $y_{MIC}$ are used for training and the remaining 40% for verification. It shows that the self-verification accuracy is higher than other models of the same type. What’s more, under MSE criteria, the feature extraction network with 30 highest-layer neurons can not only obtain the highest self-verification rate, but also show good characterizing information ability for the corresponding feature $y_{MIC}$. For these reasons, the value of $\tau$ is determined to be 9.

FIGURE 10. The optimal feature ranking under MIC processing.

Fig. 10 shows the optimal feature ranking under MIC processing. The horizontal axis represents the feature priority ranking obtained by the MIC method when the $R_{high}$ is from 20 to 60. The vertical axis represents the self-verification classification accuracy obtained by the LVQ classifier after the current feature is combined with all the features that were prioritized by itself. It can be seen that the features appeared after the accuracy has become stable have little influence on the classification accuracy. These features are the redundant information that needed to be eliminated by the MIC operation. It’s known together with Table 2, when $R_{high}$ is 30, the number of MIC features is the smallest, the model can get the best redundancy elimination effect and the highest MIC classification accuracy.

FIGURE 11. Confusion matrices comparison.

All the training samples in these methods are randomly selected. The rows correspond to the predicted class (Output Class) and the columns correspond to the true class (Target Class). The diagonal cells correspond to observations that are correctly classified. In MFE, the large number of predicted output is not placed on the master diagonal, which reflects the low characterizing information ability of the features. In comparison, the latter three methods have obvious improvements and the number of misclassification is decreasing. The detailed comparison between different methods is shown in Table 3.

Part (1) of Table 3 shows the LVQ classification accuracies under different sub-feature-extraction-methods that make up the DLVM, their values varies with different methods. The total rate is the ratio of the total correctly classified samples number and the total samples number. The number of testing samples in five stages are 208, 224, 468, 468, and 234. The WPD-AE, MIE(10), and MIE(100) can better characterize S1 and S5. The EMD-AE and TIR have a good effect on the classification of S2 and S4. The FFT and TIR obtain a relatively high accuracy for S3. But even the maximum classification accuracy of a single feature is only 50.437%, which is too low to satisfy the requirement of practical production.

Part (2) of Table 3 summarizes the performances of different classifiers and integration methods. The number
TABLE 3. Classification accuracy comparison under different methods. $S^*$ represents the number of samples that belong to stage $S^*$ and are correctly classified.

| Part | Method name | S1 | S2 | S3 | S4 | S5 | Total rate(%) (correct/total) |
|------|-------------|----|----|----|----|----|-----------------------------|
| (1) Feature extraction methods in MFE | FFT | 75 | 45 | 257 | 153 | 77 | 37.890(607/1602) |
| | WPD-AE | 87 | 69 | 207 | 180 | 90 | 39.513(633/1602) |
| | EMD-AE | 69 | 88 | 140 | 184 | 65 | 34.082(546/1602) |
| | TIR | 52 | 108 | 263 | 242 | 143 | 50.457(808/1602) |
| | MIE(10) | 105 | 163 | 207 | 185 | 108 | 47.940(768/1602) |
| | MIE(100) | 111 | 80 | 213 | 159 | 139 | 43.820(702/1602) |
| | MFE | 205 | 181 | 294 | 272 | 166 | 69.788(1118/1602) |
| (2) Classifiers and integration methods | NB [29] | 12 | 54 | 103 | 184 | 32 | 23.970(384/1602) |
| | KNN [32] | 35 | 48 | 179 | 201 | 99 | 35.081(562/1602) |
| | RUSBoost [30] | 58 | 63 | 130 | 237 | 87 | 35.892(575/1602) |
| | DNN [33] | 228 | 234 | 451 | 441 | 179 | 95.693(1533/1602) |
| | DBN30 | 213 | 217 | 467 | 441 | 227 | 97.690(1565/1602) |
| | DBN20-MIC | 210 | 212 | 410 | 426 | 247 | 93.945(1505/1602) |
| | DBN40-MIC | 220 | 226 | 431 | 407 | 230 | 94.507(1514/1602) |
| | DBN50-MIC | 239 | 216 | 408 | 449 | 224 | 95.880(1536/1602) |
| | DBN60-MIC | 226 | 209 | 431 | 465 | 222 | 96.941(1553/1602) |
| | E-BP | 206 | 198 | 438 | 460 | 233 | 95.818(1555/1602) |
| | E-PNN [31] | 214 | 155 | 468 | 453 | 248 | 96.005(1588/1602) |
| | MFE-DMIC [19] | 208 | 198 | 466 | 459 | 231 | 97.503(1562/1602) |
| | DVLM | 223 | 228 | 437 | 462 | 221 | 98.065(1571/1602) |

The MFE features has a total rate of 69.788%, but the overall classification error is still large. The DBN30 network with $R_{high} = 30$ deeply extract the MFE features, although not eliminate the redundancy, it is still effective. The DBN-MIC method (the method integrated DBN and MIC) with different $R_{high}$ reduce the number of features and obtain better results. The DVLM model with $R_{high} = 30$ and $\tau = 9$ correctly classify up to 1571 testing samples which corresponds to 98.065% of 1602 samples. The experimental results also show that the DVLM has a better classification performance as compared to BP, NB, RUSBoost, PNN, and KNN. We also compare with a latest methods DNN and MFE-DMIC. In DNN, The number of neurons in the three hidden layers is same and is 50. To obtain the optimal weights of DNN, we use the function “fminunc” in Matlab to calculate the minimum value of the cost function. The input training features of SVM and BP are $y_9^{MIC}$, here renamed them as E-BP, E-ANN, and E-SVM, E stands for evolution. The benchmark methods we used for comparison are configured with Matlab default parameters. These configurations can basically ensure a good working state of the corresponding model. The number of neurons in BP hidden layer is 20; the spread in PNN is 0.8; the values of SVM penalty parameters $c = 2$, kernel function parameter $g = 0.05$.

A well designed method is based on the principle of minimum-in-cluster-distance and maximum-between-cluster-distance. So we introduced the t-SNE method to realize the visualization of all 4004 samples by minimizing the standardized Euclidean distance between the original data and the reconstructed data. Fig. 12 shows the distributions of different dimensional features obtained by different methods when perplexity $= 200$. The distances between well-separated clusters in a t-SNE plot only reflects relative position and has no physical meaning. The dimension of features for each method is given in the caption of the subpicture below.

The features in Fig. 12(a) is obtained by using the MFE method; due to the lack of deep feature extraction, there were a lot of mutual inclusions and overlapping. For the features in Fig. 12(b), some points belonging to Stage 3 still have the sample dispersion problem. In Fig. 12(c), there is no interruption and isolation, the same stage feature points are relatively aggregated but exists over-high dimensions in feature space. Compared with others, the DVLM gets a more proper in-cluster distance and a larger between-cluster distance.

V. DISCUSSION

Faced with the difficulties in data feature extraction and low classification efficiency caused by the uncertainty and the non-uniqueness of the pressure response, the accumulation of 30 million pressure sampling points is used to develop an well-testing data classification model. The powerful database and the numerous simulation results ensured the reliability and feasibility of the proposed model.

Based on the well-testing data analysis, we find that the data streams at different stages show different time-domain characteristics due to the influence of well test workflow or geological environment, which can be abstractly expressed by the features obtained by using time-domain, frequency-domain and time-frequency domain methods.
In this way, not only the types of extracted features can be significantly enriched, but also the misjudgment caused by the artificial classification can be avoided. Therefore, MFE scheme is put forward.

As for improving the information charactering ability, we introduce four classical methods to expand the types of extracted features. We make a classification accuracy comparison of the features obtained by the sub-methods that make up the proposed method in this article. However, the single sub-method is not efficient; Table 3 Part (1) shows that the maximum classification accuracy is only 39.513%. Although the TIR achieved the highest accuracy of 50.437% among all MFE sub-methods, its low accuracy indicates that the regression method is only suitable for the case of low-dimensional and linearly separable variables in traditional well-testing data analysis, but not for the case of nonlinear variables and multi-dimensional features in well-testing stage classification. Therefore, it can be concluded that the classical methods have significant limitations in classifying well-testing data and working stage decision-making. The MFE enriched the types of data features, extracts 38 feature elements, and obtain a better accuracy of 69.788%, but it lacks the selection of major and minor elements, so the further feature mining is still needed. As for the pressure derivative method recommended by current experts, its working principle is similar to linear regression. Although we do not use pressure derivative for classification accuracy analysis, the comparison with linear regression based features such as TIR and MIE (see Table 3) can already show the superiority of MFE method.

In order for the deeper expression of the data, we use DBN to learn the MFE features. Compared with the latest model MFE-DMIC, this article is more detailed in selecting the number of neurons in DBN highest layer, and its experimental results are more convincing. We first try to analyze the impact of different \( R_{\text{high}} \) values and then assign priority ranking. As shown in Fig. 9, different types of features have different information-carrying abilities, so for features ranked before the 34th, the classification accuracy will step up as the priority increases. When the optimal number of neurons is kept within a certain range (red dots in Fig. 9), the classification accuracy remains stable. There is no clear relationship between the classification accuracy and the neurons size, not the bigger the better. Different numbers of neurons may have the same effect, but for layers that differ by only one or two neurons, the effect of the two may differ. The results in Fig. 8 and Table 2, we find that the small \( R_{\text{high}} \) is insufficient to summarize and reflect the sample laws in the training set. The large \( R_{\text{high}} \) may firmly learn and store some non-relevant information that should be loosely remembered. The classification results of DBN30 (97.690%) and other DBN integration methods in Table 3 show that non-linear and high-dimensional learning mode of the deep learning is more conducive to distinguishing and extracting the effective features.
FIGURE 13. The process of acquiring MFE features. Here we randomly select one sample for analysis. Figure 13.a shows a complete well test curve. Figure 13.b shows the obtained FFT coefficients, three IMF components of EMD, and the gradient direction of linear regression. The stage curves in Fig. 1.b are taken from Fig. 13.a. Fig. 13.c shows the obtained WPD reconstruction coefficients of the stage 1, the number of decomposed layers is 3. The number of decomposition coefficients in other stages is also 8, and the decomposition method is the same. It should be noted that the components of EMD and the coefficients of WPD shown here are only the intermediate parameters, their final features can be obtained only after the approximate entropy calculation is done. For illustrative purpose, we removed the labels from axes. For stage curves, x-label: time, y-label: pressure. For FFT, x-label: frequency, y-label: amplitude. For IMF, it represents the component of each frequency in original signals and has no unit. For LR, x-label: time, y-label: pressure. For WPD, because the reconstructed signal is the same size as the original signal, so x-label: time, y-label: pressure.
TABLE 4. Main definitions of the proposed model.

| Symbol | Quantity | Symbol | Quantity |
|---|---|---|---|
| $s(n_c^*)$ | Data in $n_c^*$ group | $n_c^*$ | Group number of samples |
| $N$ | Length of $s(n_c^*)$ | $n_l$ | Number of IMF components |
| $C_L$ | WPD decomposed level | $nd$ | Partitions divided $n_c^*$ group data into |
| $k_F$ | Number of FFT coefficients | $(d_c^F(n_c))$ | WPD coefficients of $n_c^*$ group data |
| $X(n_c^*)$ | WPD-AE of $s(n_c^*)$ | $A_Pn_c^*(n_c)$ | Approximate entropy (AE) |
| $\hat{\beta}(n_c^*)$ | FFT coefficients of $s(n_c^*)$ | $(e_c^F(n_c))$ | IMF components of $n_c^*$ group data |
| $\hat{\beta}(n_c^*)$ | EMD-AE of $s(n_c^*)$ | $r_c^F$ | Regression coefficients of TIR |
| $Reg_{(n_c^*)}$ | MIE points | $r_c^M$ | Regression coefficients of MIE |
| $reg_{(n_c^*)}$ | TIR parameters | $\theta_{opt}$ | Optimum DBN weights |
| $Attr_{train}$ | Training set features | $R_{\text{sup}}$ | Number of neurons in DBN highest layer |
| $Attr_{test}$ | Testing set features | $G_{\text{target}}$ | Goal value of LVQ |
| $y_{\text{input}}$ | Features output from $R_{\text{sup}}$ | $y_{\text{MIC}}(n_c^*)$ | MIC purified features |
| $\text{Acc}$ | Classification accuracy | $\text{Err}$ | Errors used for evaluating LVQ parameters |
| $Tag_{(n_c^*)}$ | Predicted testing tag | $Tag(n_c^*)$ | Tag corresponding to $s(n_c^*)$ |

Information representing the characteristics of the original signal. On the one hand, the MFE does obtain the effective time-frequency domain features, on the other hand, deep feature learning is really necessary. Meanwhile, from 38 to 30, the DBN reduces the number of feature elements by characterizing the low-dimensional features of sequence data in high dimensions.

Considering the feature redundancy and the correlation between the feature elements, we use the MIC to prioritize the feature elements and select the representative features. This achieves the purpose of reducing the elements of features without losing any useful information. In Table 2, the MIC self-verification accuracy appeared to be higher than the MIC classification accuracy. It is mainly because the former trained network, with the same training and testing sets, produces the network parameters with better adaptability to training the samples. The comparison between the two verifies the effectiveness of the designed DBN network. Fig. 10 supplements the rationality of choosing $R_{\text{high}} = 30$ by giving the ranking of the number of effective feature elements after MIC processing. In Fig. 10, the feature output by DBN30 network has the least number of elements and has the highest classification efficiency. By analyzing the correlation between different feature elements in DBN30, MIC purifies the redundancy and obtains fewer (reduced the number of dimensions from 30 to 9) but more representative features (improved the classification accuracy from 97.690% to 98.065%, see Table 3), which provide an efficient feature input for the LVQ classifier.

As for improving the generalization ability, the LVQ classifier is proven effective in finding a nonlinear mapping between oil test data and operation stage. Usually, the geological data can lead to a bias in interpreting the accuracies and errors. To reduce the system error and improve the classification efficiency, we use the average CV error to complete the parameter optimization and obtain the optimal number of neurons in the LVQ competition layer. By introducing the cumulative error $e$ and the LVQ parameters, we provide a calculation formula of evaluation error $\text{Err}$. As its applications, Fig. 7 analyzes the impact of different LVQ parameters, provides the distribution of average error and minimum error, and determines $R_{\text{CL}} = 30$, $G_{\text{target}} = 0.28$. For any input of unknown stage, the predicted tags can be automatically and accurately output. It can be seen from Fig. 4 that the structure of LVQ can not only classify the feature, but also further achieve the feature reconstruction and deep training process. The comparison results in Table 3 with different classifiers, deep neural network and the latest MFE-DMIC method verify the superiority of LVQ.

All in all, creating an ensemble data classifier from a set of classic data processing methods is found promising in enhancing the classification performance. The classifier can exploit capabilities of different individual classical methods. The proposed DVLM model is a suitable option in the well-testing analysis for determining the correct well-testing working stage, providing a reliable reference for oil production management.

The proposed model also exists some limitations and shortcomings. First, the feature types are not enough and comprehensive. Second, the sample size for deep learning and network training is not large. Third, the proposed model, which integrates DBN and LVQ, achieves good classification performance, but only considers the DBN highest layer optimization. So, in future work, three “more” should be studied in-depth. More feasible methods should be used to maximize the data information mining. More field samples including more complex operations should be collected to realize the better training of the learning network. More influencing factors should be considered in the adjustment and the optimization of model structure.
VI. CONCLUSION
Aiming at improving the oil well-testing stage classification accuracy, we firstly explain the processing principle and the construction process of the proposed model in light of the complexity, randomness and regional characteristics of data. Secondly, we emphasize the importance of the diversity of features, DBN neurons size, feature priority and LVQ parameters. Then the features extracted by the MFE method are used as the input of the structure-optimized DBN network, and combined with the MIC and improved LVQ classifier, the high-precision stage classification is realized. The experimental results show that, benefiting from the improvement in information characterizing ability of the features and the generalized classification ability of the classifier, the DVLIM has higher accuracy of 98.065% than the classical mathematical methods, various integration methods, and the latest deep learning methods. The proposed model reduces human intervention in the stage classification process and has broad application prospects in stochastic and non-scheduled data processing.

REFERENCES
[1] H. Rahmanifar and T. Płaskinska, “Application of artificial intelligence techniques in the petroleum industry: A review,” Artif. Intell. Rev., vol. 52, no. 4, pp. 2295–2318, Dec. 2019.
[2] D. A. Wood and B. B. Hazra, “Characterization of organic-rich shales for petroleum exploration & exploitation: A review—Part 3: Applied geomechanics, petrophysics and reservoir modeling,” J. Earth Sci., vol. 28, no. 6, pp. 49–73, Jun. 2017.
[3] K.-R. Qian, Z.-L. He, and Y.-Y. Chen, “Intelligent prediction and integral analysis of shale oil and gas sweet spots,” Petroleum Sci., vol. 15, no. 4, pp. 744–755, Nov. 2018.
[4] Z. Ge, Z. Song, S. X. Ding, and B. Huang, “Data mining and analytics in the process industry: The role of machine learning,” IEEE Access, vol. 5, pp. 20590–20616, 2017.
[5] L. Chun, X. X. Chunhua, Y. Jian, and B. Junliang, “A method for coastal works,” J. Mech. Eng., vol. 15, no. 4, pp. 1–17, Jan. 2017.
[6] P. P. Roy, G. Zhong, and M. Cheriet, “Tandem hidden Markov models using deep belief networks for offline handwriting recognition,” Frontiers Inf. Technol. Electron. Eng., vol. 18, no. 7, pp. 978–988, Jul. 2017.
[7] S.-Y. Shao, W.-J. Sun, R.-Q. Yan, P. Wang, and R. X. Gao, “A deep learning approach for fault diagnosis of induction motors in manufacturing,” Chin. J. Mech. Eng., vol. 30, no. 6, pp. 1347–1356, Nov. 2017.
[8] D.-L. Feng, M.-Q. Xiao, X.-Y. Liu, H.-F. Song, Z. Yang, and Z.-W. Hu, “Finite-sensor fault-diagnosis simulation study of gas turbine engine using information entropy and deep belief networks,” Frontiers Inf. Technol. Electron. Eng., vol. 17, no. 12, pp. 1287–1304, Dec. 2016.
[9] M. Hammad, P. Prawiak, K. Wang, and U. R. Acharya, “ResNet-attention model for human authentication using ECG signals,” Expert Syst., vol. e12547, pp. 1–17, Mar. 2020.
[10] A. Sabri Shahrabakabi, A. S. Imran, N. Olifati, and T. Svendsen, “A comparative study on CMD and wavelet transform methods for ECG signal classification,” Circuits, Signal Process., vol. 38, no. 8, pp. 3501–3520, Aug. 2019.
[11] M. Hammad, Y. Liu, and K. Wang, “Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint,” IEEE Access, vol. 7, pp. 26527–26542, 2019.
[12] J. K. Kim, Y. S. Han, and J. S. Lee, “Particle swarm optimization-deep belief network based random classifier prediction model for highly class imbalance problem,” Concurrency Comput., Pract. Exper., vol. 29, no. 11, Jun. 2017, Art. no. e4128.
[13] Q. Zou, L. Ni, T. Zhang, and Q. Wang, “Deep learning based feature selection for remote sensing scene classification,” IEEE Geosci. Remote Sens. Lett., vol. 12, no. 11, pp. 2321–2325, Nov. 2015.
[14] P. Yu and Z. Tang, “Application of Bayesian theory to predict permeability in reservoir,” J. Cent. South Univ., vol. 46, no. 3, pp. 970–982, 2015.
[15] M. H. Kesikoglu, U. H. Atasever, C. Ozkan, and E. Besdok, “The usage of rusboost boosting method for classification of impervious surfaces,” ISPRS Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci., vols. XLII–B7, pp. 981–985, Oct. 2016.
[16] W. J. Al-Mudhafar, “Integrating lithofacies and well logging data into smooth generalized additive model for improved permeability estimation: Zubair formation, South Rumaila oil field,” Mar. Geophys. Res., vol. 40, no. 1, pp. 315–332, Jan. 2017.
[17] Y. Li, H. Wang, M. Wang, P. Lian, T. Duan, and B. Ji, “Automatic identification of carbonate sedimentary facies based on PCA and KNN using logs,” Well Logging Technol., vol. 41, no. 1, pp. 57–61, Feb. 2017.
[18] M. Shahriari, D. Pardo, A. Picon, A. Galdran, J. Del Ser, and C. Torres-Verdín, “A deep learning approach to the inversion of borehole resistivity measurements,” Comput. Geosci., vol. 24, no. 3, pp. 971–994, Jun. 2020.
[19] T. Kohonen, M. Schroeder, and T. Huang, Self-Organizing Maps. Berlin, Germany: Springer, 2001.
[20] A. Ghosh, M. Biehl, and B. Hammer, “Performance analysis of LVQ algorithms: A statistical physics approach,” Neural Netw., vol. 19, nos. 6–7, pp. 817–829, Jul. 2006.
[21] L. C. Lanzarini, A. V. Monte, A. F. Bariviera, and P. J. Santana, “Simplifying credit scoring rules using LVQ + PSO,” Kybernetes, vol. 46, no. 1, pp. 8–16, Jan. 2017.
[37] M. Li, F. Chen, M. Lei, and C. Li, “Near-infrared spectrum of coal origin identification based on LVQ with SVM algorithm,” Spectrosc. Spectral Anal., vol. 36, no. 9, pp. 2793–2797, 2016.
[38] E. Saeedi, Y. Kong, and M. S. Hossain, “Side-channel attacks and learning-vector quantization,” Frontiers Inf. Technol. Electron. Eng., vol. 18, no. 4, pp. 511–518, Apr. 2017.
[39] X. Liu, D. Mao, Z. Wang, and Y. Yang, “Rapid identification of Coix seed varieties by near infrared spectroscopy,” Spectrosc. Spect. Anal., vol. 34, no. 5, pp. 1259–1263, 2014.
[40] Y. Cui, J. Jiang, Z. Lai, W. Wong, and Z. Hu, “An integrated optimisation algorithm for feature extraction, dictionary learning and classification,” Neurocomputing, vol. 275, pp. 2740–2751, Jan. 2018.
[41] G. Sun, J. Li, J. Dai, Z. Song, and F. Lang, “Feature selection for IoT based on maximal information coefficient,” Future Gener. Comput. Syst., vol. 89, pp. 606–616, Dec. 2018.
[42] J. C. Tung, S. P. Sastry, and S. M. Shontz, “Performance characterization of nonlinear optimization methods for mesh quality improvement,” Eng. Comput., vol. 28, no. 3, pp. 269–286, Jul. 2012.
[43] A. Alhawarat and Z. Salleh, “Modification of nonlinear conjugate gradient method with weak wolfe-powell line search,” Abstract Appl. Anal., vol. 2017, pp. 1–6, Mar. 2017.
[44] K. Nordhausen, “An introduction to statistical learning-with applications in R by Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani,” Int. Statist. Rev., vol. 82, no. 1, pp. 156–157, 2014.
[45] S. Arlot and A. Celisse, “A survey of cross-validation procedures for model selection,” Statist. Surv., vol. 4, pp. 40–79, Jul. 2010.
[46] The Concept and Partition Principle of Training Set, Verification Set and Testing Set (Chinese), vol. 32, no. 2, pp. 113–139, Sep. 2010.

XIN FENG was born in Xinzhou, Shanxi, China. He is currently pursuing the Ph.D. degree in instrument science and technology with Tianjin University, Tianjin, China. His research interests include well logging, deep learning, pattern recognition, process monitoring and fault diagnosis, and their applications. His work has achieved the First Prize in competition at 2018 CPIPC and 2019 RoboWork.

SHAOHUI LI was born in 1983. He received the Ph.D. degree in measuring and controlling technologies and instruments from Tianjin University, Tianjin, China, in 2012. He is currently a Senior Engineer with the Tianjin Research Institute for Water Transport Engineering, Ministry of Transport, Tianjin. His research interests include intelligent drilling tool and multi-parameter data measuring and processing.

XINGWEI HOU was born in Shijiazhuang, Hebei, China. He is currently pursuing the Ph.D. degree in instrument science and technology with Tianjin University, Tianjin, China. His research interests include near infrared spectroscopy, embedded system design, and signal processing.

MENGQIU ZHANG was born in Tai’an, Shandong, China. She is currently pursuing the Ph.D. degree in biomedical engineering with Tianjin University, Tianjin, China. Her research interests include signal detecting and processing, sensor engineering, neural networks, and signal analysis.