Petroleum Science 20 (2023) 1411–1428

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

A deep kernel method for lithofacies identification using conventional well logs

Shao-Qun Dong [a, b], Zhao-Hui Zhong [a, b], Xue-Hui Cui [a, b], Lian-Bo Zeng [a, c, *], Xu Yang [a, b], Jian-Jun Liu [a, b], Yan-Ming Sun [a, b], Jing-Ru Hao [a, b]

[a] *State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing, 102249, China*
[b] *College of Science, China University of Petroleum, Beijing, 102249, China*
[c] *College of Geoscience, China University of Petroleum, Beijing, 102249, China*

ARTICLE INFO

Article history:
Received 20 April 2022
Received in revised form
29 July 2022
Accepted 30 November 2022
Available online 5 December 2022

Edited by Jie Hao and Teng Zhu

Keywords:
Lithofacies identification
Deep kernel method
Well logs
Residual unit
Kernel principal component analysis
Gradient-free optimization

ABSTRACT

How to fit a properly nonlinear classification model from conventional well logs to lithofacies is a key problem for machine learning methods. Kernel methods (e.g., KFD, SVM, MSVM) are effective attempts to solve this issue due to abilities of handling nonlinear features by kernel functions. Deep mining of log features indicating lithofacies still needs to be improved for kernel methods. Hence, this work employs deep neural networks to enhance the kernel principal component analysis (KPCA) method and proposes a deep kernel method (DKM) for lithofacies identification using well logs. DKM includes a feature extractor and a classifier. The feature extractor consists of a series of KPCA models arranged according to residual network structure. A gradient-free optimization method is introduced to automatically optimize parameters and structure in DKM, which can avoid complex tuning of parameters in models. To test the validation of the proposed DKM for lithofacies identification, an open-sourced dataset with seven conventional logs (GR, CAL, AC, DEN, CNL, LLD, and LLS) and lithofacies labels from the Daniudi Gas Field in China is used. There are eight lithofacies, namely clastic rocks (pebbly, coarse, medium, and fine sandstone, siltstone, mudstone), coal, and carbonate rocks. The comparisons between DKM and three commonly used kernel methods (KFD, SVM, MSVM) show that (1) DKM (85.7%) outperforms SVM (77%), KFD (79.5%), and MSVM (82.8%) in accuracy of lithofacies identification; (2) DKM is about twice faster than the multi-kernel method (MSVM) with good accuracy. The blind well test in Well D13 indicates that compared with the other three methods DKM improves about 24% in accuracy, 35% in precision, 41% in recall, and 40% in F1 score, respectively. In general, DKM is an effective method for complex lithofacies identification. This work also discussed the optimal structure and classifier for DKM. Experimental results show that $(m_1, m_2, 0)$ is the optimal model structure and linear SVM is the optimal classifier. $(m_1, m_2, 0)$ means there are m_1 KPCAs, and then m_2 residual units. A workflow to determine an optimal classifier in DKM for lithofacies identification is proposed, too.

Abbreviations

The abbreviations used in this article are summarized in Table 1 below.

1. Introduction

Lithofacies identification is important in reservoir characterization since it is an important prerequisite for subsequent sedimentary facies analysis and reservoir modeling (Dong et al., 2016; Shen et al., 2019). Well logs are practical data for lithofacies identification because lithofacies usually have direct or indirect influences on the signals of geophysical logs. However, porosity, permeability, oil and gas bearing properties, bedding, and fractures all affect logging responses besides lithology (Mansouri-Daneshvar et al., 2015; H Wang et al., 2021; Dong et al., 2022b). Hence,

[*] Corresponding author. State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing, 102249, China.
E-mail address: lbzeng@sina.com (L.-B. Zeng).

https://doi.org/10.1016/j.petsci.2022.11.027
1995-8226/

Table 1
Term abbreviations used in this work.

Full name	Abbr.	Full name	Abbr.
Deep kernel method	DKM	Randomized coordinate shrinking	RACOS
Principal component analysis	PCA	Soft version of max	Softmax
Kernel PCA	KPCA	K-nearest neighbor	KNN
Support vector machines	SVM	Linear SVM	linear-SVM
Kernel Fisher discriminant	KFD	Radial basis function SVM	RBF-SVM
Multi kernel SVM	MSVM	Random forest	RF
Multi KFD	MKFD	Gradient boosting decision tree	GBDT
Linear discriminant analysis	LDA	Naive Bayes	NB

lithofacies identification is a typical nonlinear classification problem, especially in the case of complex lithology, and nonlinear features extraction and use are quite significant for lithofacies identification. Nowadays, machine learning methods have drawn wide attention in addressing the problem of lithofacies identification (Liu et al., 2020b; Sun et al., 2020). Among them, kernel methods are a kind of methods that employ kernel functions to extract and use nonlinear features.

For a nonlinear classification problem using conventional logs, it is linearly inseparable in the original space (Phillips and Abdulla, 2021). However, a proper nonlinear mapping into a higher-dimensional nonlinear feature space can make this problem linearly separable according to the Cover's theorem (Cover, 1965). The key in the process of raising dimension lies in determining the nonlinear mapping function. Unfortunately, it is quite difficult to obtain a proper explicit nonlinear mapping function (Yuan et al., 2019). Nevertheless, to address the determination problem of explicit mapping function, kernel methods use the kernel function to indirectly calculate the inner product between all pairs of data samples in the high-dimensional feature space to avoid explicit calculation of nonlinear feature mapping, which is called “kernel trick” (Dong et al., 2016, 2020a, 2020c; Shi et al., 2020). Since their introduction in the mid-1990s, kernel methods have been proven successful for many machine learning problems (Yan et al., 2012; QS Wang et al., 2021).

Commonly used kernel methods can be divided into single-kernel methods (e.g., support vector machines, SVM; relevance vector machine, RVM; kernel Fisher discriminant, KFD; kernel principal component analysis, KPCA) and multi-kernel methods (e.g., multi SVM, MSVM; multi-kernel RVM, MKRVM; multi KFD, MKFD).

(1) Applications in lithofacies identification by single-kernel methods

SVM maps original samples into a higher dimensional feature space by kernel function, and a hyperplane classifier is built to identify lithofacies (Tharwat and Hassanien, 2019b; Al-Najjar and Pradhan, 2021; Chen et al., 2021; Farouk et al., 2021). It shows superiority during the disposal of small samples, nonlinearity, and high-dimensional classification problems (Liu et al., 2020b; Xiao et al., 2020). For the crystalline rocks from Chinese Continental Scientific Drilling Main Hole (CCSD-MH), the classification accuracy of SVM is about 5% higher than that of back propagation neural network (BPNN) (Deng et al., 2017). Different from SVM, RVM introduces a Bayesian approach that provides a posterior probability output (Zhang et al., 2017; Mohebian et al., 2018; Zhu et al., 2021), which avoids determination of penalty parameter and needs few vectors (Liu et al., 2020a). RVM is as accurate as SVM, but RVM is faster than SVM (Tipping, 2001; Liu et al., 2020a).

Improved SVM algorithms are widely used in lithofacies identification as well, such as particle swarm optimization SVM (PSO-SVM), proximal SVM (PSVM), and least square SVM (LS-SVM). PSO-SVM is an SVM method where its parameters are optimized by PSO algorithm (Qu et al., 2020; Li et al., 2021; Xu et al., 2021). Considering the complexity of volcanic reservoirs and the influence of alterations in the Songliao Basin of China, the PSO-SVM method was used for lithological identification and obtained a high accuracy (Pan et al., 2022). Different from SVM, PSVM constructs parallel planes to approximate data classes, which has the advantages of reducing memory usage and computing time (Malik and Mishra, 2016). PSVM was used to distinguish between limestone and shale obtained by Barnett shale gas. Compared with the standard SVM, PSVM can be better used for 3D seismic lithofacies classification (Zhao et al., 2014). LS-SVM is a special form of SVM that considers equality type constraints (Suykens and Vandewalle, 1999). LS-SVM algorithm is suitable for solving nonlinear problems with small samples, and the speed of LS-SVM is faster than that of SVM (Liu et al., 2021).

As a kernel method of dimension reduction, KFD can map samples in a high-dimensional feature space into a low-dimensional feature space with the largest category difference for classification by kernel trick (Shi et al., 2020). In the lithology identification of complex lithology in Junggar Basin of China, KFD obtained an over 3% improvement compared with QDA (quadratic discriminant analysis), which is a common nonlinear method of linear discriminant analysis (LDA) (Dong et al., 2016).

Besides classification and regression, kernel methods can also be used to extract features. Kernel PCA is an improved principal component analysis (PCA) by kernel tricks (Anowar et al., 2021; TH Wang et al., 2021; Yu et al., 2021). KPCA is promising in decoupling the nonlinear correlation of variables (well logs) (Ge et al., 2014; Heidary, 2015).

Single-kernel methods are commonly used. They will choose an optimal nonlinear mapping by determining optimal kernel parameters. How to choose proper kernel function and set kernel parameters are important since they can significantly influence stability, generalization, and accuracy of kernel methods (Ortiz-García et al., 2009; Lin et al., 2017; Tharwat, 2019).

(2) Applications in lithofacies identification by multi-kernel methods

The nature of the feature space corresponding to a single kernel decides some deficiencies of single-kernel methods, such as instability (Lanckriet et al., 2004). The diversity of logging parameters against lithology exacerbates instability of single-kernel methods in classification problems (Lan et al., 2021). To handle the aforementioned problems, multi-kernel methods, which combine effective features from different feature spaces, are introduced to enhance the stability and accuracy of categorical effects (Gu and Liu, 2016). In recent years, more and more researches related to multi-kernel learning have been published, such as the extension of SVM, RVM, KFD, etc. (Areiza-Laverde et al., 2020).

MSVM integrates the nonlinear models of SVM in different feature spaces, which makes it more flexible and stable than a single-kernel SVM model (He et al., 2016). For the facies classification in the Hugoton and Panoma gas fields of USA, MKRVM was proposed and showed a better prediction performance than SVM and RVM. SVM and RVM had similar accuracy, while MKRVM showed about 2% improvement compared with them (Liu et al., 2020a). MKFD can fully extract useful information in different feature spaces through multiple KFD (Dong et al., 2022a). For the carbonate reservoirs of Zagros Basin in Iraq, the lithofacies identification experiments show MKFD outperforms KFD with an over 10% increase in accuracy, precision, and recall (Dong et al., 2022b). Note that MSVM has a better performance than single-kernel SVM, but the computing speed is relatively slow (Tang et al., 2019). Time efficiency is a common issue for multi-kernel methods.

The reviews above mentioned indicating that the ability of an algorithm to extract nonlinear features is significant to improve the accuracy of lithofacies recognition. In many cases, multi-kernel methods can obtain a better lithofacies identification, especially in complex reservoirs. This indicates a powerful ability of extracting nonlinear log features for distinguishing lithofacies is still needed. Therefore, this work proposes a novel method named as deep kernel method (DKM) to deeply mine and utilize nonlinear features for identifying lithofacies, which aims to improve kernel methods by artificial neural networks. DKM uses a residual neural network module to implement deep learning; employs a set of KPCA models to extract nonlinear features; manipulates the extracted nonlinear features to predict lithofacies labels by a classifier. In order to verify the effectiveness of this method in lithofacies identification, a series of comparative experiments were carried out using DKM and auxiliary kernel methods (KFD, SVM, MSVM). In addition, the optimal structure and classifier of DKM will be discussed, too.

2. Principle of deep kernel method (DKM) for lithofacies identification

2.1. DKM model for lithofacies identification based on ResNet structure and KPCA

Conventional well logs labelled by lithofacies descriptions will be randomly divided into training (e.g., 80%) and test (e.g., 20%) data as shown in Fig. 1a and b. The training data is used to train a DKM model while the test data is used to validate the prediction ability of the built model.

DKM consists of two parts, namely a feature extractor (Fig. 1c) and a classifier (Fig. 1d).

The feature extractor is composed of a series of kernel principal component analysis (KPCA) models, which are m_1 KPCA models, m_2 KPCA models with residual units, and m_3 KPCA models in sequence. This feature extractor aims to extract nonlinear features which help distinguish different lithofacies by nonlinear mapping in different scaled feature spaces. Compared with the ResNet, KPCA is used to extract features by kernel tricks instead of full fully connected layer or convolution layer since KPCA is more suitable for dealing with nonlinear feature extraction. There is a skip connection (Orange arrow line in Fig. 1c) for each KPCA model in the residual unit structure. For the $i-$ th (between 2nd and m_2- th) KPCA model in the residual unit structure, the input is a combination of the outputs of $(i-1)-$ th and $(i-2)-$ th KPCA models rather than only the output of $(i-1)-$ th KPCA. The skip connection can ensure an effective deep feature extractor even though there are failed KPCA models (Zhang et al., 2021). This deep feature extraction structure allows information to flow between layers, preserving information useful for accurate classification and preventing information loss due to dimensional changes. The residual unit structure is expressed by Eq. (1).

$$H(\boldsymbol{X})=\boldsymbol{X}+F(\boldsymbol{X};\sigma_i) \tag{1}$$

where $\boldsymbol{X}$ is the input; $F(\boldsymbol{X};\sigma_i)$ is the kernel feature extractor KPCA; $H(\boldsymbol{X})$ is the output of each residual unit and the input of the next residual unit; σ_i is the parameter in the $i-$ th kernel.

In this work, the classifier can be any classifier rather than softmax classifier in ResNet. In artificial neural networks, the error back propagation, chain rule and gradient-based optimization are common ways to optimize parameters in networks. However, these will not work in DKM since the component of DKM is non-derivable. Hence, a gradient-free optimization method (randomized coordinate shrinking, RACOS) is introduced to handle this problem, which is presented in Section 2.4. In the parameter optimization of DKM, negative accuracy is adopted as loss function. The parameters to optimize include the numbers of KPCA (m_1, m_2, and m_3), the feature dimension of each layer, the parameters in classifier, and parameter σ_i in each KPCA model. According to the principle of minimum loss, the parameters will be updated iteratively to make the DKM model with a minimum loss.

2.2. Feature extractor (KPCA) in DKM

The main idea of KPCA is to project original data into high-dimensional feature space through nonlinear mapping, and then perform linear dimensionality reduction in the high-dimensional space. It is proven to be an effective method with a good theoretical foundation to handle nonlinear classification problems.

By nonlinear mapping $\boldsymbol{\Phi}$, each sample in the low dimensional input space $\boldsymbol{X}$ will be mapped as a sample in the high-dimensional feature space ($\mathbf{R}^q$), where $\boldsymbol{X}\ =(\boldsymbol{X}_1,\boldsymbol{X}_2,...,\boldsymbol{X}_p)(\boldsymbol{X}_i\in\mathbf{R}^n,i=1,2,...,p)$.

$$\boldsymbol{\Phi}:\begin{cases}\mathbf{R}^n\rightarrow\mathbf{R}^q\ q\geq n\\ \boldsymbol{X}\rightarrow\boldsymbol{\Phi}(\boldsymbol{X})\end{cases} \tag{2}$$

As shown in Fig. 2, the combination of a proper nonlinear mapping by kernel functions and dimension reduction by PCA can make data separable. This makes KPCA able to work in the feature extractor of DKM.

In KPCA, the data has been centralized, which means that the data has zero mean, and the covariance matrix can be calculated by Eq. (3) (Schölkopf et al., 1998).

$$\boldsymbol{C}=\frac{1}{p}\sum_{j=1}^{p}\boldsymbol{\Phi}(\boldsymbol{X}_j)\boldsymbol{\Phi}^T(\boldsymbol{X}_j) \tag{3}$$

where p is the number of samples. The eigenvector formula is constructed to find non-zero eigenvalue λ and eigenvector $\boldsymbol{v}\in\boldsymbol{Y}$ as displayed in Eq. (4).

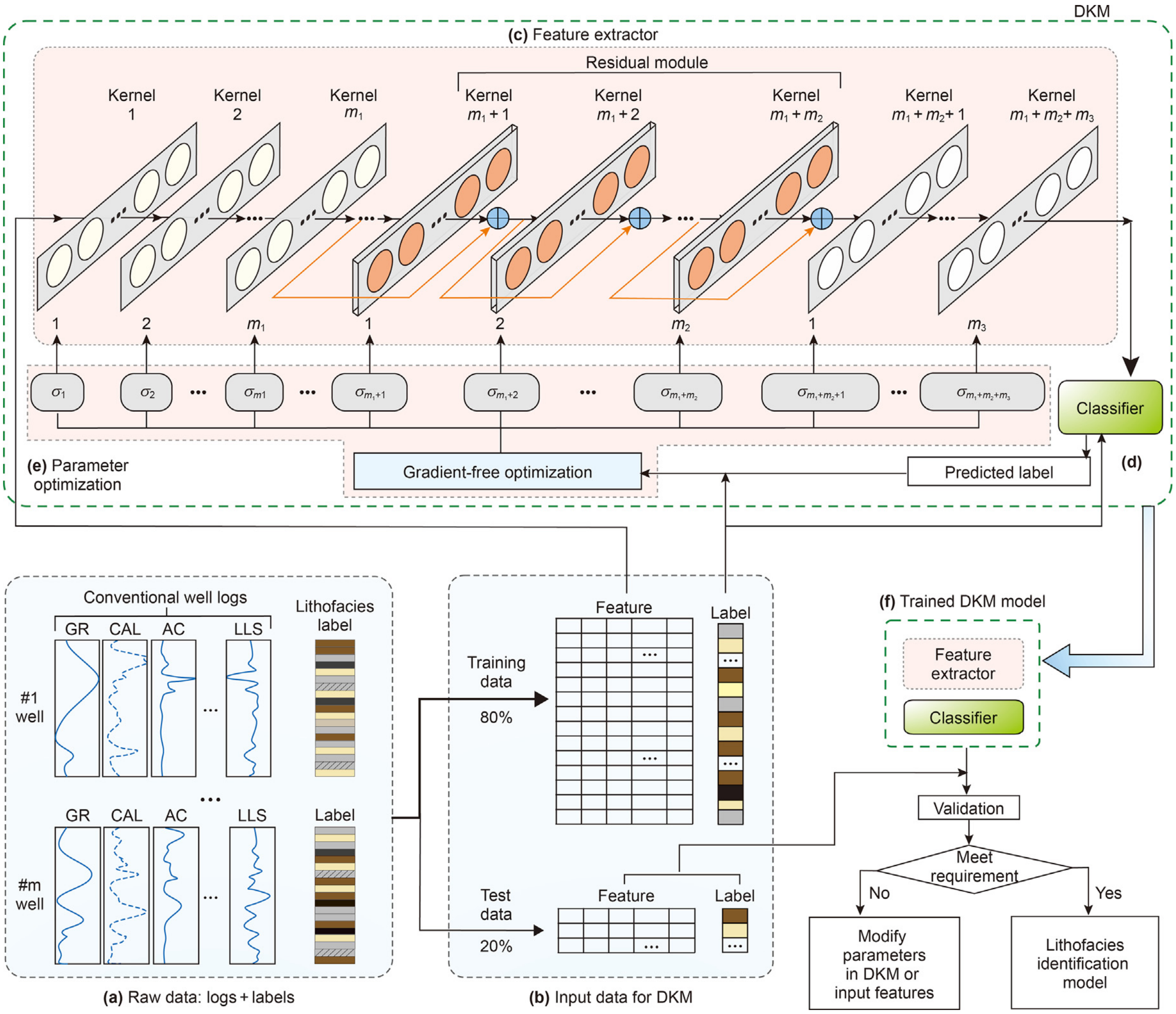


Fig. 1. Schematic diagram of the deep kernel method (DKM) for lithofacies identification.

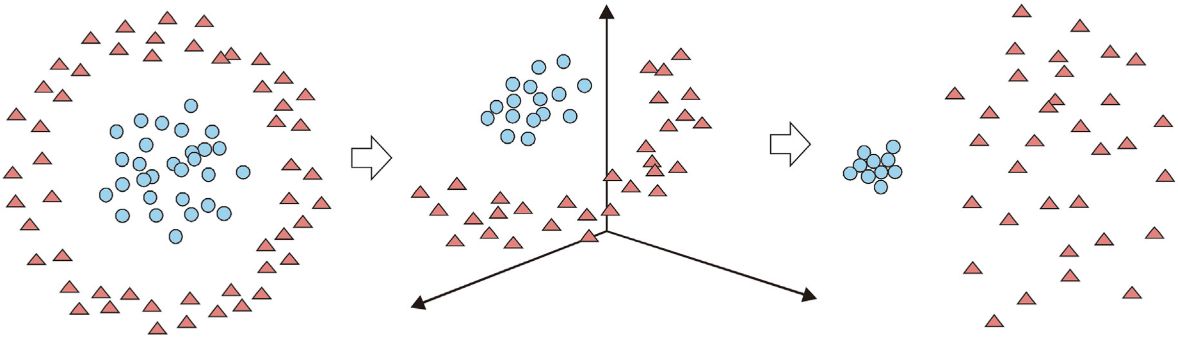


Fig. 2. Schematic diagram of KPCA.

$$\lambda \boldsymbol{v} = \boldsymbol{C}\boldsymbol{v} \tag{4}$$

According to Mercer's condition (Mercer, 1909), the following kernel function can be obtained:

$$k(\boldsymbol{X}_i, \boldsymbol{X}_j) = \boldsymbol{\Phi}^T(\boldsymbol{X}_i)\boldsymbol{\Phi}(\boldsymbol{X}_j) = \boldsymbol{\Phi}(\boldsymbol{X}_i)\cdot\boldsymbol{\Phi}(\boldsymbol{X}_j) \tag{5}$$

Because any feature vector in space $\boldsymbol{Y}$ can be linearly represented by all samples in the space, namely

$$\boldsymbol{v} = \sum_{i=1}^{p} \alpha_i \boldsymbol{\Phi}(\boldsymbol{X}_i) \tag{6}$$

Then Eq. (4) can be transformed into

$$\lambda \sum_{i=1}^{p} \alpha_i \boldsymbol{\Phi}(\boldsymbol{X}_i) = \frac{1}{p} \sum_{j=1}^{p} \boldsymbol{\Phi}(\boldsymbol{X}_j) \boldsymbol{\Phi}^T(\boldsymbol{X}_j) \sum_{i=1}^{p} \alpha_i \boldsymbol{\Phi}(\boldsymbol{X}_i) \tag{7}$$

Multiply $\boldsymbol{\Phi}^T(\boldsymbol{X}_k)$ in both sides of Eq. (7), Eq. (8) can be obtained. Combine Eq. (3) with Eq. (8), and then Eq. (9) will be obtained.

$$\lambda \sum_{i=1}^{p} \alpha_i \boldsymbol{\Phi}^T(\boldsymbol{X}_k) \boldsymbol{\Phi}(\boldsymbol{X}_i) = \frac{1}{p} \sum_{i=1}^{p} \left(\boldsymbol{\Phi}^T(\boldsymbol{X}_k) \left(\sum_{j=1}^{p} \boldsymbol{\Phi}(\boldsymbol{X}_j) \boldsymbol{\Phi}^T(\boldsymbol{X}_j) \right) \alpha_i \boldsymbol{\Phi}(\boldsymbol{X}_i) \right) \tag{8}$$

$$\lambda \boldsymbol{K} \boldsymbol{\alpha} = \frac{1}{p} \boldsymbol{K}^2 \boldsymbol{\alpha} \tag{9}$$

where $\boldsymbol{K}$ is the kernel function matrix of $p \times p$ dimension; $\boldsymbol{\alpha}$ is the column vector containing $\alpha_1, \alpha_2, \ldots, \alpha_p$. Eq. (9) can be further simplified as

$$p \lambda \boldsymbol{\alpha} = \boldsymbol{K} \boldsymbol{\alpha} \tag{10}$$

Then λ and $\boldsymbol{\alpha}$ can be determined by solving Eq. (10).

For a new sample $\boldsymbol{X}_j (j = 1, 2, \ldots, p)$, its principal component after nonlinear dimension reduction will be

$$\boldsymbol{v}^T \cdot \boldsymbol{\Phi}(\boldsymbol{X}_j) = \sum_{i=1}^{p} \alpha_i \left[\boldsymbol{\Phi}(\boldsymbol{X}_i)^T \cdot \boldsymbol{\Phi}(\boldsymbol{X}_j) \right] \tag{11}$$

In this process, the determination of nonlinear mapping is a difficult problem. However, KPCA cleverly employs kernel function to implement a nonlinear mapping and avoids determining the explicit nonlinear mapping function (Esmaeili et al., 2020). At present, Gaussian kernel function is the most commonly used kernel function, as shown below.

$$k(\boldsymbol{X}_i, \boldsymbol{X}_j) = exp\left(-\sigma \|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2 \right) \tag{12}$$

where σ is the parameter that determines how much the input variable is scaled in a learning algorithm.

Replace the inner product with a kernel function, then Eq. (11) becomes

$$\boldsymbol{v}^T \cdot \boldsymbol{\Phi}(\boldsymbol{X}_j) = \sum_{i=1}^{p} \alpha_i k(\boldsymbol{X}_i, \boldsymbol{X}_j) \tag{13}$$

By default, $\boldsymbol{\Phi}(\boldsymbol{X})$ has been centralized in the above derivation process. If it is not centralized, diagonalized $\boldsymbol{K}_c$ should be used to replace $\boldsymbol{K}$ for the above solution process (Li et al., 2020). The expression for $\boldsymbol{K}_c$ is

$$\boldsymbol{K}_c = \boldsymbol{K} - \boldsymbol{l}_p \boldsymbol{K} - \boldsymbol{K} \boldsymbol{l}_p + \boldsymbol{l}_p \boldsymbol{K} \boldsymbol{l}_p \tag{14}$$

where $\boldsymbol{l}_p$ is the $p \times p$ matrix, and each of its elements is $\frac{1}{p}$.

2.3. Classifier in DKM

In this work, eight representative methods in Fig. 3 are selected as candidate classifiers for DKM, namely soft version of max (Softmax), conventional machine learning methods (linear discriminant analysis, LDA; Naive Bayes, NB; K-nearest neighbor, KNN; linear SVM, linear-SVM), kernel method (radial basis function SVM, RBF-SVM) and ensemble learning methods (random forest, RF; gradient boosting decision tree, GBDT). The performances of the eight classifiers in DKM will be compared. A workflow to determine an optimal classifier will be provided in Section 4.3.

(1) Softmax maps the output of multiple neurons to the interval of (0,1), which makes a multi-classification transform into a probability comparison problem;
(2) LDA aims to maximize the inter-class mean and minimize the intra-class variance. This means that the data is projected on a low dimension, the projection points of the same type of data are as close as possible, and the center points of the projection points of different types of data are as far as possible;
(3) NB is one of the top 10 data mining algorithms (Zhang et al., 2020). It is a classification method based on Bayes' theorem and characteristic condition independence hypothesis. For a given training set, the joint probability distribution of input and output is first learned based on the independent assumption of feature conditions. Then, for a given input $\boldsymbol{x}$, the output $\boldsymbol{y}$ with the maximum posteriori probability is obtained based on this model using Bayes' theorem;
(4) KNN only determines the classification of the samples to be divided according to the category of the nearest one or several samples. KNN has the advantages of simplicity and small recognition error (Pablo et al., 2016);
(5) The basic idea of SVM classification is to use maximum interval for classification (Yin and Yin, 2016; Hou et al., 2020). The application condition of linear-SVM is that when samples are linearly separable in the feature space, a hyperplane can be found that can completely divide samples of different classes. The application condition of RBF-SVM is when the sample is nonlinearly separable in the feature space. The data is mapped to the higher dimensional feature space by kernel function to make the sample linearly separable in the higher dimensional space, and then the classification hyperplane is constructed (Tambe et al., 2018);
(6) RF is an ensemble algorithm composed of many decision trees based on Bagging idea (Qiao and Chang, 2021; Zhou et al., 2022). Sub-datasets are generated by random sampling with replacement and random features selection, and the predicted result on the strength of voting mechanism (Xu and Sun, 2018). On account of the above basic ideas, RF is more robust to noise and has better generalization ability (Yates and Islam, 2021);

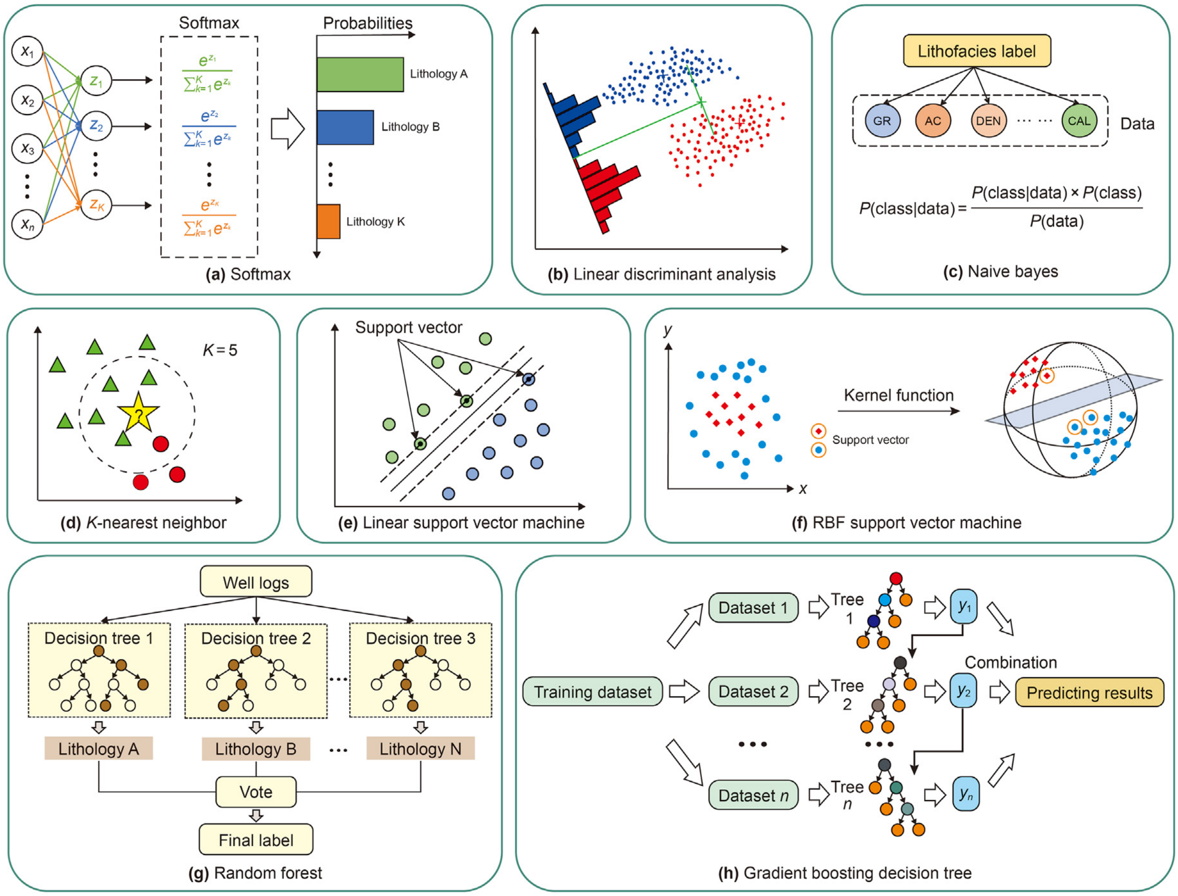


Fig. 3. Schematic diagram of candidate classifiers for DKM.

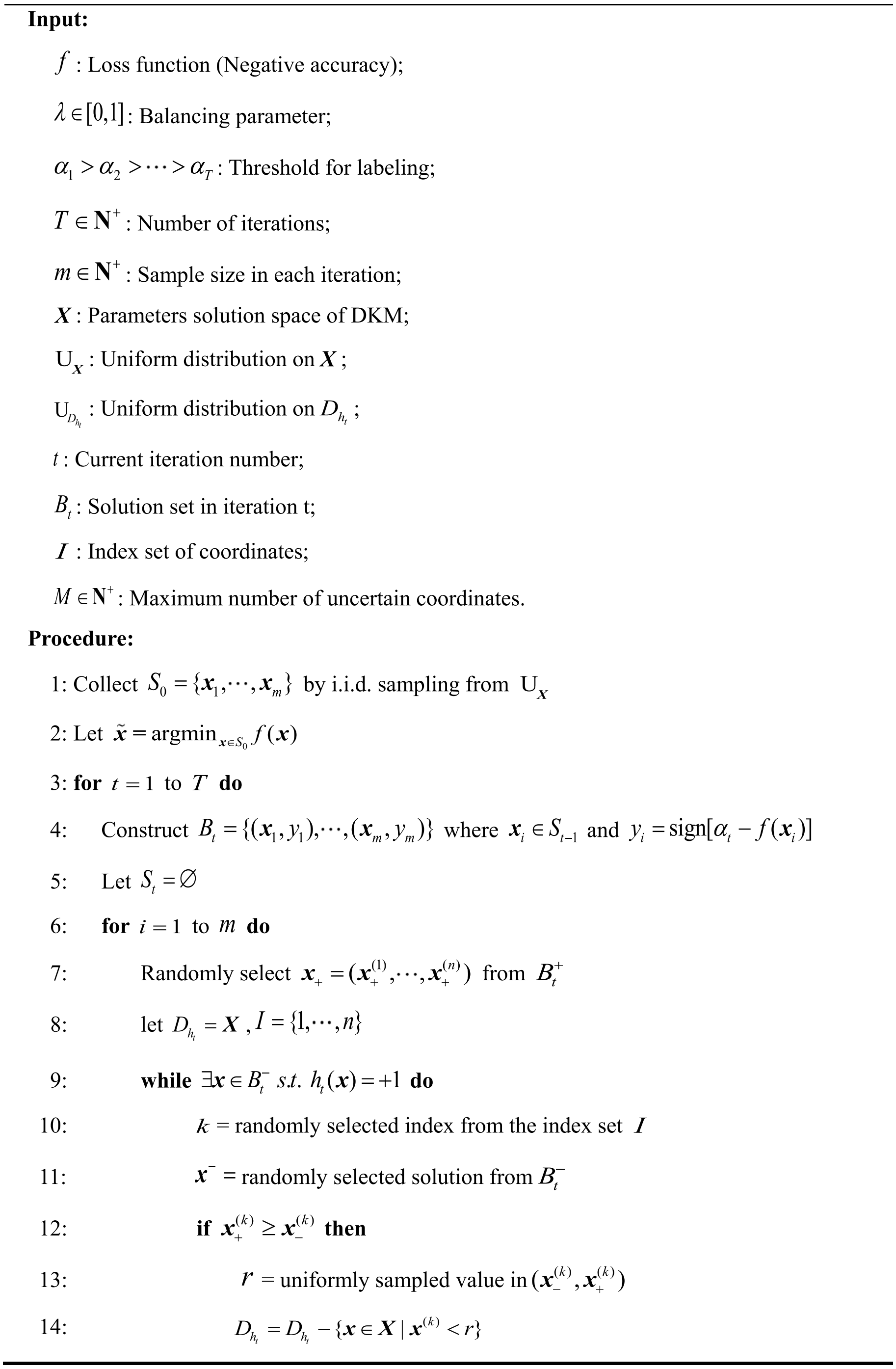
(7) GBDT integrates multiple weak classifiers into strong classifiers using the boosting approach (Liang et al., 2020). GBDT has been widely used in many fields due to its strong interpretation and good application effect.

2.4. Parameter optimization in DKM

Gradient-based optimization algorithms (SGD, Adam, etc.) can effectively solve most neural network models which support the success of deep learning (Wu et al., 2021). Due to the basis of gradient, the model is required to be differentiable. However, the proposed DKM model is a generalization of a neural network and the characteristics of differentiable may not be guaranteed. Hence, a gradient optimization method will not be the first choice since it may only obtain local optimum. Gradient-free optimization method based on "smart" sampling is a good choice since it does not depend on derivatives and can reach a global optimization (Liu et al., 2015). The fewer requirements on the nature of the problem make it more suitable for DKM. In this work, a global gradient-free optimization algorithm, called randomized coordinate shrinking (RACOS) (Yu et al., 2016), will be employed to determine optimal parameters in kernels, classifier, and layer structure. RACOS is good at addressing high-dimensional complex optimization problems. Its optimal parameter selection range keeps the shape of hypercube, which enables it to quickly find a suitable solution in the high-dimensional searching space.

Algorithm 1 is the process of optimization algorithm in DKM. $\boldsymbol{X}$ is the solution space of DKM parameters. The initial solution set is sampled from $\mathrm{U}_{\boldsymbol{X}}$, namely $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$, and the current optimal solution $\tilde{\boldsymbol{x}}$ is obtained. The global optimal solution is found through T times of iterative updating. Each cycle in T contains m times resampling. Firstly, m samples in solution set are divided into positive and negative samples by function sign$[\alpha_t - f(\boldsymbol{x}_i)]$ according to the selected threshold α_t. The positive samples are expressed as $h_t(\boldsymbol{x}) = +1$, and the negative samples are expressed as $h_t(\boldsymbol{x}) = -1$. Then m samples will be resampled. A classifier h_t needs to be found to make the positive region $D_{h_t} = \{\boldsymbol{x} \in \boldsymbol{X} | h_t(\boldsymbol{x}) = +1\}$ induced by it as close to region $D_{\alpha_t} = \{\boldsymbol{x} \in \boldsymbol{X} | f(\boldsymbol{x}) \le \alpha_t\}$ through continuous learning. The new samples are generated by sampling from the uniform distribution $\mathrm{U}_{D_{h_t}}$ on the positive class region D_{h_t} with probability λ and from the uniform distribution $\mathrm{U}_{\boldsymbol{X}}$ on the whole solution space $\boldsymbol{X}$ with probability $1 - \lambda$ through the process of Sampling (h_t, λ). Finally, the samples obtained by resampling will be recorded, and the best DKM parameters will be found.

Algorithm 1. Optimization algorithm in DKM.



```
Input:
  $f$ : Loss function (Negative accuracy);
  $\lambda \in [0,1]$ : Balancing parameter;
  $\alpha_1 > \alpha_2 > \cdots > \alpha_T$ : Threshold for labeling;
  $T \in \mathbf{N}^+$ : Number of iterations;
  $m \in \mathbf{N}^+$ : Sample size in each iteration;
  $\boldsymbol{X}$ : Parameters solution space of DKM;
  $\mathrm{U}_{\boldsymbol{X}}$ : Uniform distribution on $\boldsymbol{X}$ ;
  $\mathrm{U}_{D_{h_t}}$ : Uniform distribution on $D_{h_t}$ ;
  $t$ : Current iteration number;
  $B_t$ : Solution set in iteration t;
  $I$ : Index set of coordinates;
  $M \in \mathbf{N}^+$ : Maximum number of uncertain coordinates.
Procedure:
  Collect $S_0 = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$ by i.i.d. sampling from $\mathrm{U}_{\boldsymbol{X}}$
  Let $\tilde{\boldsymbol{x}} = \mathrm{argmin}_{\boldsymbol{x} \in S_0} f(\boldsymbol{x})$
  for $t = 1$ to $T$ do
      Construct $B_t = \{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_m, y_m)\}$ where $\boldsymbol{x}_i \in S_{t-1}$ and $y_i = \mathrm{sign}[\alpha_t - f(\boldsymbol{x}_i)]$
      Let $S_t = \varnothing$
      for $i = 1$ to $m$ do
          Randomly select $\boldsymbol{x}_+ = (\boldsymbol{x}_+^{(1)}, \cdots, \boldsymbol{x}_+^{(n)})$ from $B_t^+$
          let $D_{h_t} = \boldsymbol{X}$ , $I = \{1, \cdots, n\}$
          while $\exists \boldsymbol{x} \in B_t^-$ s.t. $h_t(\boldsymbol{x}) = +1$ do
              $k$ = randomly selected index from the index set $I$
              $\boldsymbol{x}^-$ = randomly selected solution from $B_t^-$
              if $\boldsymbol{x}_+^{(k)} \geq \boldsymbol{x}_-^{(k)}$ then
                  $r$ = uniformly sampled value in $(\boldsymbol{x}_-^{(k)}, \boldsymbol{x}_+^{(k)})$
                  $D_{h_t} = D_{h_t} - \{\boldsymbol{x} \in \boldsymbol{X} \mid \boldsymbol{x}^{(k)} < r\}$
```

```
15:             else
16:                 $r$ = uniformly sampled value in $(\boldsymbol{x}_+^{(k)}, \boldsymbol{x}_-^{(k)})$
17:                 $D_{h_t} = D_{h_t} - \{\boldsymbol{x} \in \boldsymbol{X} \mid \boldsymbol{x}^{(k)} > r\}$
18:             end if
19:         end while
20:         while $\#I > M$ do
21:             $k$ = randomly selected index from the index set $I$
22:             $D_{h_t} = D_{h_t} - \{\boldsymbol{x} \in \boldsymbol{X} \mid \boldsymbol{x}^{(k)} \neq \boldsymbol{x}_+^{(k)}\}$, $I = I - \{k\}$
23:         end while
24:         return $h_t$
25:         $\boldsymbol{x}_i = \text{Sampling}(h_t, \lambda) = \begin{cases} \mathrm{U}_{D_{h_t}}, \text{with probability } \lambda; \\ \mathrm{U}_{\boldsymbol{X}}, \text{with probability } 1-\lambda. \end{cases}$
26:         let $S_t = S_t \cup \{\boldsymbol{x}_i\}$
27:     end for
28:     $\tilde{\boldsymbol{x}} = \text{argmin}_{\boldsymbol{x} \in S_t \cup \{\tilde{\boldsymbol{x}}\}} f(\boldsymbol{x})$
29: end for
30: return $\tilde{\boldsymbol{x}}$ and $f(\tilde{x})$
```

(***continued***).

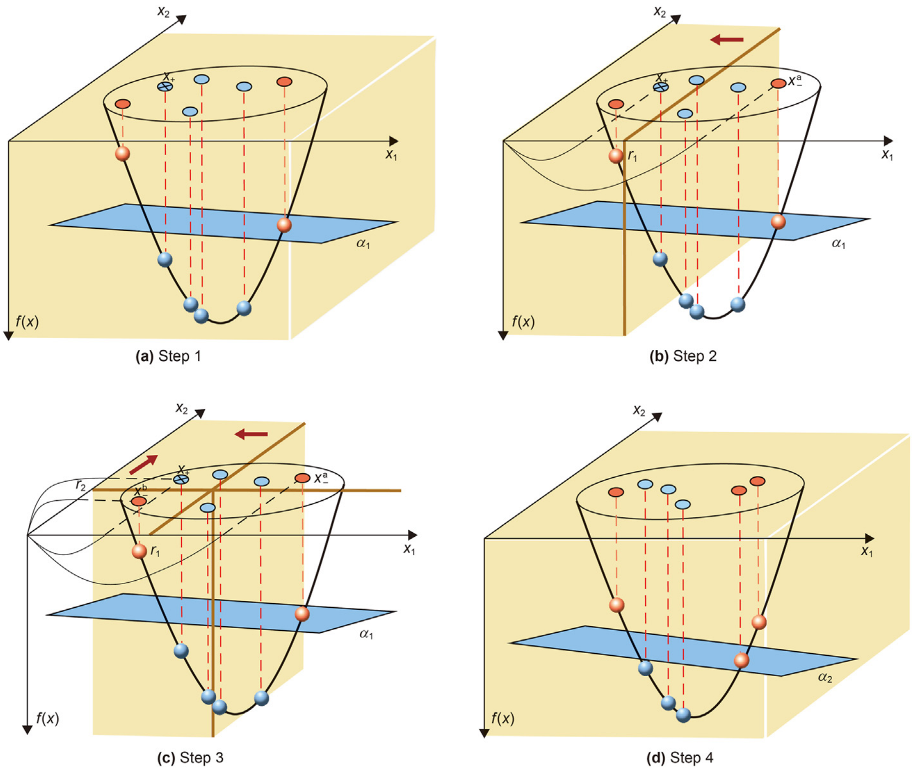
The classifier h_t satisfying the above is not unique. RACOS provides an algorithm, referring to lines 7–24 in Algorithm 1, where B_t^+ and B_t^- are the positive and negative sample set divided by the previously selected threshold α_t. In DKM, only the continuous case is considered. For the number of model layers in DKM parameters, the method of rounding after optimization is adopted. The schematic diagram of two-dimensional continuous process is shown in Fig. 4, where $m = 6$ is selected and the yellow area on the $\boldsymbol{x}$ plane represents D_{h_t}. Six sample points are sampled from the whole space and the current optimal solution is recorded. Then the current optimal solution needs to be constantly updated after T iterations to find the global optimal solution. The first step is shown in Fig. 4a, the threshold α_1 is selected to divide the samples into four positive samples and two negative samples. B_t^+ and B_t^- are positive and negative sample set, respectively. Then find a classifier h_t, and make the positive region $D_{h_t} = \{\boldsymbol{x} \in \boldsymbol{X} | h_t(\boldsymbol{x}) = +1\}$ induced by it as close as possible to region $D_{\alpha_t} = \{\boldsymbol{x} \in \boldsymbol{X} | f(\boldsymbol{x}) \leq \alpha_t\}$. Initially, D_{h_t} is the entire space and $\boldsymbol{x}_+$ is randomly selected from the positive samples. Since $\exists \boldsymbol{x} \in B_t^- s.t. h_t(\boldsymbol{x}) = +1$, so enter the while loop for the second step, as shown in Fig. 4b. Here, $k = 1$ and $\boldsymbol{x}_- = \boldsymbol{x}_-^a$ are randomly selected. Since $\boldsymbol{x}_+^{(1)} \leq \boldsymbol{x}_-^{a(1)}$, r_1 is randomly selected in the interval $(\boldsymbol{x}_+^{(1)}, \boldsymbol{x}_-^{a(1)})$, so the positive sample region is $D_{h_t} = D_{h_t} - \{\boldsymbol{x} \in \boldsymbol{X} | \boldsymbol{x}^{(1)} > r_1\}$. Since $\boldsymbol{x} \in B_t^- s.t. h_t(\boldsymbol{x}) = +1$ still exists, step 3 is performed, as shown in Fig. 4c. Select $k = 2$ and $\boldsymbol{x}_- = \boldsymbol{x}_-^b$ at random. Since $\boldsymbol{x}_+^{(2)} \geq \boldsymbol{x}_-^{b(2)}$, r_2 is randomly selected in the interval $(\boldsymbol{x}_-^{b(2)}, \boldsymbol{x}_+^{(2)})$, so the positive sample region is $D_{h_t} = D_{h_t} - \{\boldsymbol{x} \in \boldsymbol{X} | \boldsymbol{x}^{(2)} < r_2\}$. The next step is the resampling process, by sampling (h_t, λ), $\boldsymbol{x}_1$ can be resampled from the positive sample region with a probability of λ and from the whole space with a probability of $1 - \lambda$. After updating a sample point, repeat steps 2 to 3 until all six sample points are updated. The above is an iteration process. Step 4 is to reselect the threshold, as shown in Fig. 4d. Ensure that $\alpha_1 > \alpha_2$, and then repeat the above iteration process. Throughout the process, the best sample points to date will be recorded. RACOS algorithms are constantly learning and shrinking. Sampling from any region in a high-dimensional space is not easy. The RACOS algorithm's shrinking region is always a hyperrectangle, making sampling straightforward and efficient.

2.5. Workflow of lithofacies identification by DKM

The working principle of DKM is shown in Fig. 5, where Nc is the number of iterations. $\{\boldsymbol{x}_i, y_i\}$ is the input data, where $\boldsymbol{x}_i$ is a vector composed of conventional logging curves, and y_i is lithofacies label code. Then all log data is standardized by $(\boldsymbol{x}_{ij} - \text{mean}(\boldsymbol{x}_i))/\text{std}(\boldsymbol{x}_i)$, where $\boldsymbol{x}_{ij}$ is the $i -$ th log of $j -$ th sample, $\text{mean}(\boldsymbol{x}_i)$ and $\text{std}(\boldsymbol{x}_i)$ are the average value and standard deviation of the $i -$ th log. Note deep and shallow latero logs are standardized after logarithm transforming. Then, the data is divided into blind well data, training set, and test set, and then the model is trained with the training set. If DKM performs well in test set and blind well, it can be applied to lithology prediction of new wells.

3. Lithofacies identification in the Daniudi Gas Field (DGF) of Ordos Basin, China

In this work, all methods are programmed by Python and all lithofacies identifications and calculations are conducted on an Intel (R) Core (TM) computer with 2.4 GHz CPU and 64 GB of RAM.



(a) Step 1

(b) Step 2

(c) Step 3

(d) Step 4

Fig. 4. Schematic diagram of two-dimensional continuous process.

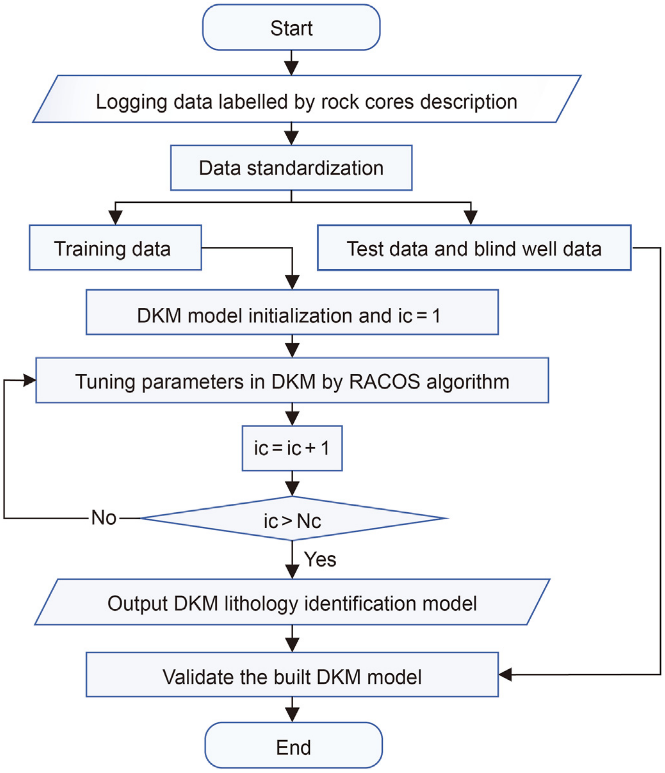

Start
Logging data labelled by rock cores description
Data standardization
Training data
Test data and blind well data
DKM model initialization and ic = 1
Tuning parameters in DKM by RACOS algorithm
ic = ic + 1
No
ic > Nc
Yes
Output DKM lithology identification model
Validate the built DKM model
End

Fig. 5. Flow chart of lithofacies identification by DKM.

3.1. Geologic settings and dataset used

An open-sourced dataset from the Daniudi Gas Field, China (Xie et al., 2018) is selected to test the proposed method for lithofacies identification. The Daniudi Gas Field is located in the Yishan slope of Ordos Basin as shown in Fig. 6a. The Yishan slope lies in the east of asymmetric syncline, which is the main body of the Ordos Basin. It is the focus of oil and gas exploration and development in the basin. The target formation is the Upper Paleozoic reservoirs, which are of low porosity and permeability and formed in a fluvial-deltaic depositional environment. There are clastic rocks, coal (C), and carbonate rock (CR). Based on the detrital grain size classification of the oil industry standard of China, the clastic rocks can be subdivided into six kinds of lithofacies, namely pebbly sandstone (PS) > 1 mm, coarse sandstone (CS) 0.5–1 mm, medium sandstone (MS) 0.25–0.5 mm, fine sandstone (FS) 0.01–0.25 mm, siltstone (S) 0.005–0.05 mm, mudstone (M) < 0.005 mm (Xie et al., 2018).

In this work, seven wells are used, namely D2, D4, D6, D9, D13, D17, and D23. For each well, seven conventional logs (GR, CAL, AC, DEN, CNL, LLD, and LLS) are measured as shown in Fig. 6b. The lithofacies labels in different depths of each well are given based on core analysis reports provided by Huabei branch of Sinopec Group (Xie et al., 2018). The sections of D2, D4, D6, D9, D13, D17, and D23 with core descriptions are 208 m, 156 m, 421 m, 51 m, 337 m, 180 m, and 123 m, respectively. In total, there are 915 samples in the dataset used. 80% of the samples will be randomly selected to be as training data to build a DKM model for lithofacies identification while the other 20% will be as the test data to test the validation of the built model. Meanwhile, the section between 2682 m and 2850 m of Well D13 will be used for blind well test.

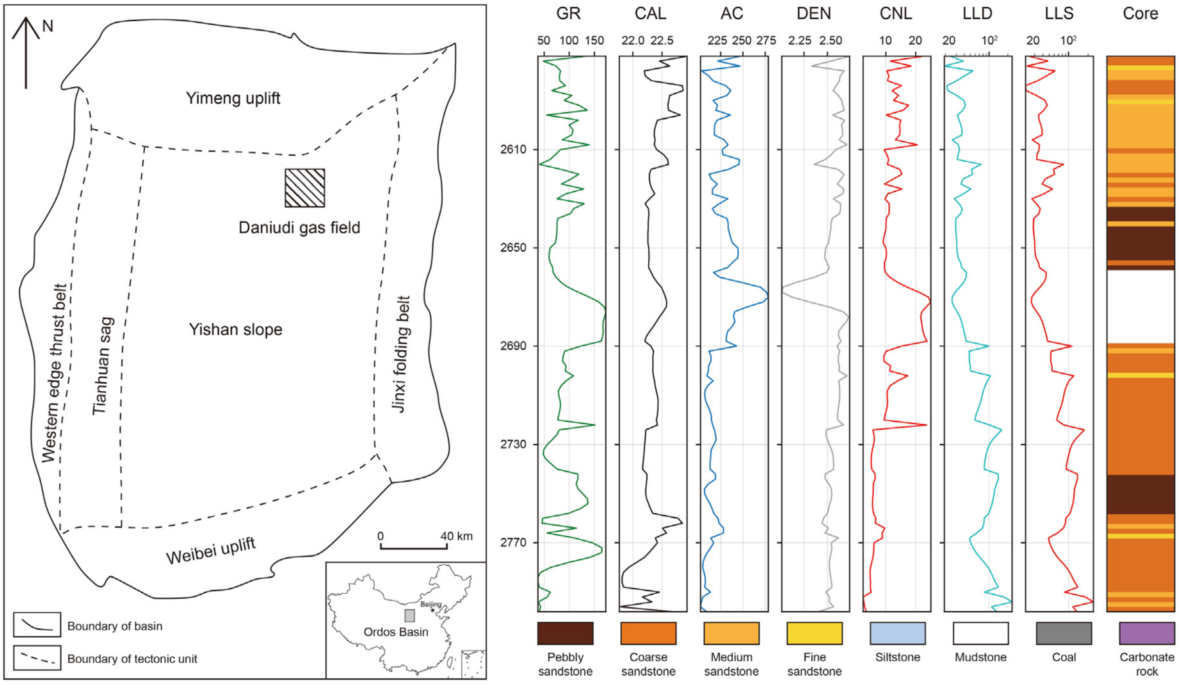


Fig. 6. Location of the study area (from Dong et al., 2020b) and a well profile with logs and lithofacies labels.

3.2. Well log responses of lithofacies

The range of well log data in the study area is shown in Table 2. The mean and standard deviation represent average values and fluctuations in the data, respectively. The minimum and maximum values correspond to the upper and lower limit of value, respectively. 25% quantile, median, and 75% quantile are the corresponding quartiles in the cumulative probability distribution of log data, which can reflect the centralized distribution and central tendency of data. If the median value is closer to 25% quantile than to 75% quantile, the values tend to be near small values and left-skewed, and vice versa.

Cross plots of well logs against lithofacies in the study area are shown in Fig. 7. Probability density curves of logs against different lithofacies are shown on the up and right sides of the cross plots. Probability density curves are obtained by kernel density method. The separations among one probability density curve of one well log and others indicate that this lithofacies is easy to be distinguished from others (Dong et al., 2022b). For example, the density of coal is lowest compared with other lithofacies, which makes AC high and DEN low. As shown on the right side of Fig. 7a, the peak of AC corresponding to coal is obviously separated from the others, which indicates AC is suitable for identifying coal from other lithofacies. Similarly, DEN of coal formation is lower than other lithofacies. On the upper side of Fig. 7c, the peak corresponding to coal is also far from the other peaks of lithofacies. Due to joints in coals, CNL values of coal is commonly high as shown on the upper side of Fig. 7c. While drilling, the CAL will be enlarged to some extent as shown on the upper side of Fig. 7b due to relatively loose structure of the coal formation. Besides coal usually has the characteristics of high resistivity as shown in Fig. 7d. Hence, these logs can be used to predict coal.

For carbonate rocks, characteristics of low CNL and high DEN can be observed in Fig. 7c. Besides characteristics of low AC can be observed on the right side of Fig. 7a, too.

For clastic rocks, mudstone is different from other clastic rocks. Typically, it has the characteristics of high GR and high CNL. The small grain size of mudstone makes radioactive elements absorbed in mudstone which subsequently leads to high GR values. CNL measures the hydrogen content index in rocks. Bound water in mudstone makes it high CNL values. For other clastic rocks, even though there are some separations in GR as shown on the upper side of Fig. 7a, there are still a lot of overlaps among these probability density curves of each lithofacies. Identifying clastic rocks with different grain sizes is important for reservoir characterization, analyses of sedimentary facies, and evaluation of fracture density. It is necessary to distinguish these clastic rocks (PS, CS, MS, FS, S). However, the overlaps make this prediction problem a complex nonlinear issue. Hence, extraction and utilization of nonlinear log features against of lithofacies is a difficult problem to solve for the lithofacies identification in the study area.

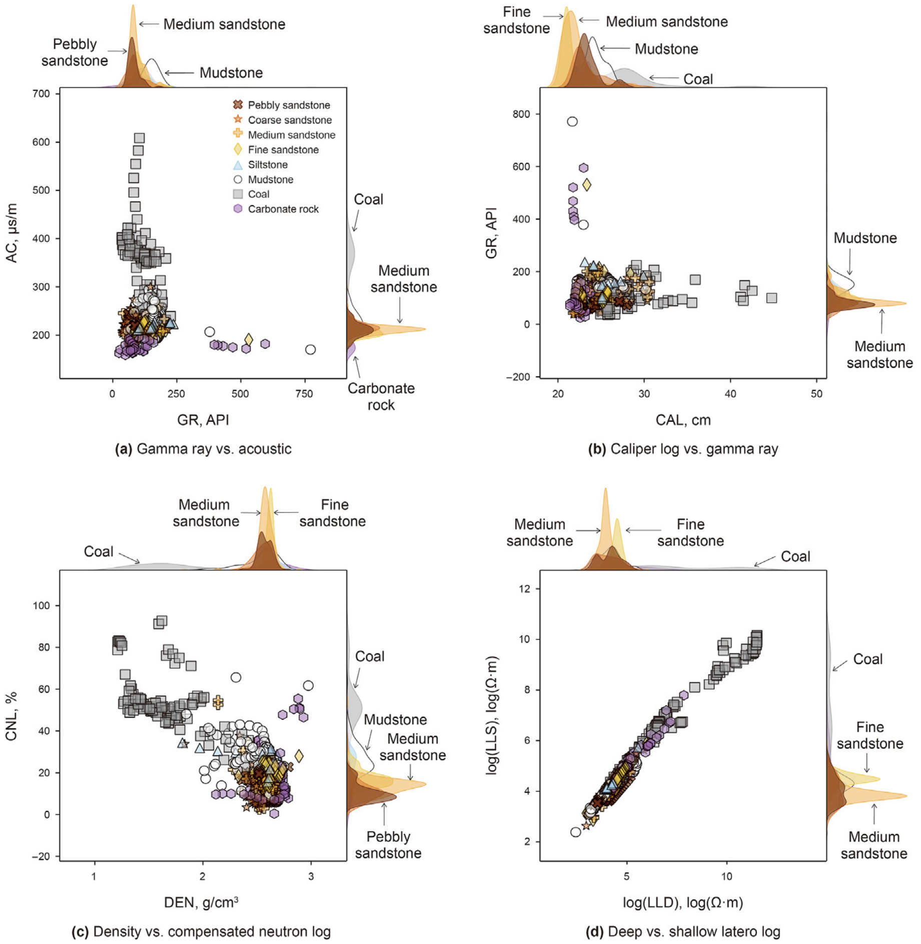
Table 2
Statistical characteristics of conventional logging data in the study area.

Variables	GR	CAL	AC	DEN	CNL	log (LLD)	log (LLS)
Unit	API	cm	μs/m	g/cm^3	%	log(Ω·m)	log(Ω·m)
Mean	109.6	24.2	229.9	2.48	20.6	1.99	1.96
Std	55.1	2.5	54.8	0.32	14.7	0.63	0.54
Min	24.3	21.4	159.0	1.21	0.4	1.06	1.03
25% quantile	78.0	22.7	204.4	2.51	10.9	1.69	1.66
Median	95.1	23.5	213.8	2.58	15.8	1.85	1.84
75% quantile	133.1	25.0	228.1	2.63	24.4	2.03	2.01
Max	771.3	44.8	608.6	2.97	92.8	5.00	4.42

3.3. DKM model for lithofacies identification

A general framework of DKM is provided in Fig. 1. Typically, m_3 will be set small even as zero as discussed in Section 4.2. Hence, m_3 is set as zeros here. The analyses in Section 4.3 show that linear-SVM is better than other classifiers. Therefore, linear-SVM was selected as the optimal classifier for DKM.

The parameters used in the DKM model for lithofacies identification are shown in Table 3, where m_1 represents the number of



(a) Gamma ray vs. acoustic

(b) Caliper log vs. gamma ray

(c) Density vs. compensated neutron log

(d) Deep vs. shallow latero log

Fig. 7. Cross plots of well logs against lithofacies labels. (b)–(d) share the same legend with (a).

Table 3
Optimization of parameters in DKM.

Tuned parameters	Search range	Optimal parameters setting
n_1	100–500	340
n_2	500–1000	978
m_1	0–10	2
m_2	0–10	4
SVM-kernel	linear	linear
C	0–150	137
KPCA-kernel	rbf	rbf

KPCA layers without residual structure, n_1 represents the feature dimension of the first $m_1 - 1$ layers of KPCA layers without residual structure, m_2 represents the number of KPCA layers with residual structure, n_2 represents the feature dimension of KPCA with residual structure and the m_1 – th layer of KPCA layers without residual structure, C represents the penalty coefficient in the SVM model, SVM-kernel represents the kernel function in the SVM model, and KPCA-kernel represents the kernel function in the KPCA module.

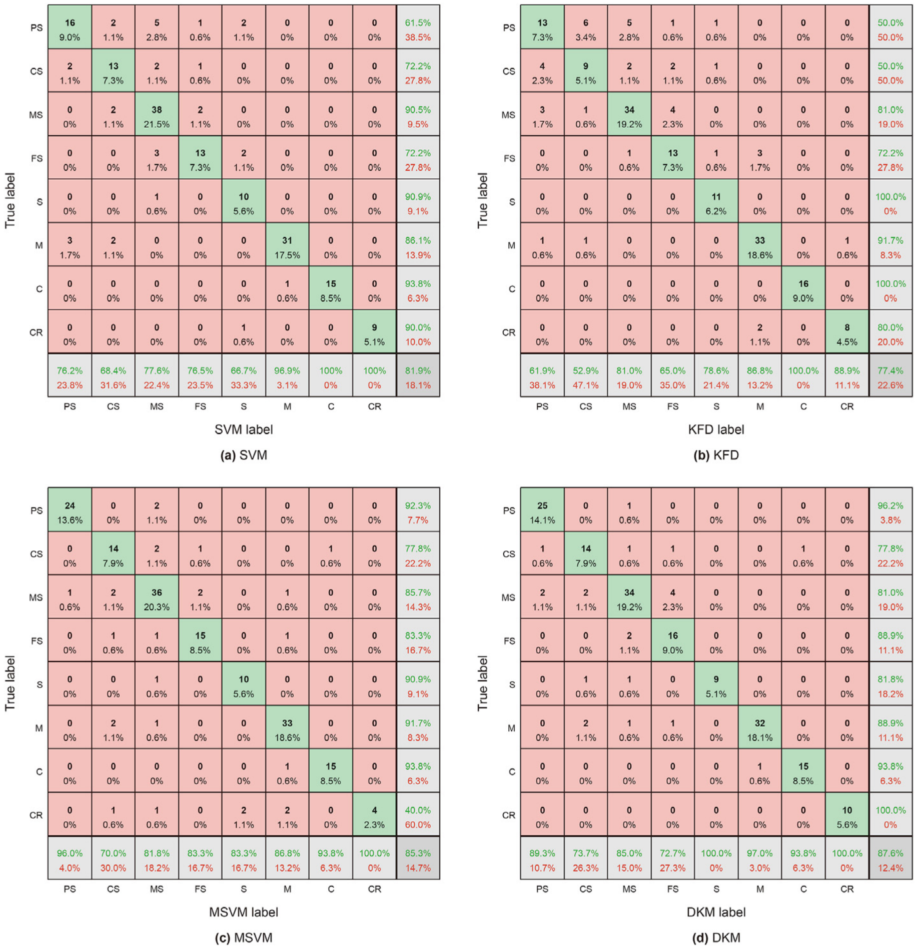
Iterative optimization of kernel parameter σ in Eq. (12) is required in DKM, and the search range is set within [0,0.3]. The optimal parameter σ of each kernel in DKM are shown in Table 4. All parameters with search range mentioned in Tables 3 and 4 were optimized by RACOS optimization algorithm. Compared with deep neural network, DKM uses a gradient-free optimization method to determine parameters automatically, which solves the problem of parameter tuning due to a large number of parameters.

Table 4
The optimal parameter σ of each kernel in DKM.

No.	1	2	3	4	5	6
σ	0.27	0.27	0.26	0.25	0.26	0.30

3.4. Comparisons of identification by DKM and other kernel methods

The confusion matrices of SVM, KFD, MSVM, and DKM for the test data are displayed in Fig. 8, where the ordinate is the real label category and the abscissa is the predicted label category. Green squares and text indicate correctly classified results, and red squares and text indicate incorrectly classified results. Accuracy is a common and intuitive evaluation index. Generally, the higher the accuracy, the better the classifier. The formula is expressed as (TP + TN)/(TP + TN + FP + FN), where TP, FP, TN, and FN respectively represent the number of positive samples correctly predicted, the number of negative samples incorrectly predicted, the number of negative samples correctly predicted and the number of positive samples incorrectly predicted. TP + TN is the sum of the numbers in the green squares, TP + TN + FP + FN is the total number of tests, so accuracy means the proportion of all correctly predicted samples in the classification model to the total test samples (XX Wang et al., 2021). The accuracy with error lines of different models is shown in Fig. 9, which is the result of random training 20 times for each model. The height of the column in the figure represents the



(a) SVM

(b) KFD

(c) MSVM

(d) DKM

Fig. 8. Confusion matrices by SVM, KFD, MSVM, and DKM. The labels are pebbly sandstone (PS), coarse sandstone (CS), medium sandstone (MS), fine sandstone (FS), siltstone (S), mudstone (M), coal (C), and carbonate rock (CR).

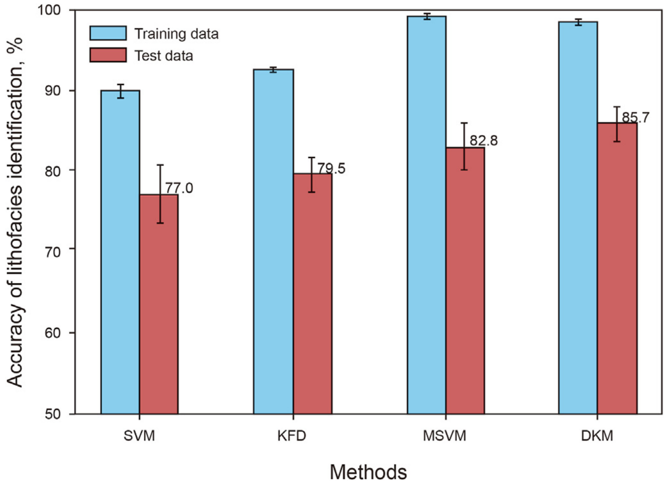


Fig. 9. Accuracy of lithofacies identification by different kernel methods. The bar height and error line represent the mean accuracy and standard deviation of 20 repeats, respectively.

Table 5
Training time of optimal models.

Model	Time
SVM	2′8″
KFD	2′2″
MSVM	8′2″
DKM	4′5″

average accuracy of the test set, and the length of the error line represents the fluctuation of the mean square error of the test set. It can be seen that the average accuracy of SVM, KFD, MSVM, and DKM is 77%, 79.5%, 82.8%, and 85.7%, respectively. The difference between training accuracy and testing accuracy can reflect the generalization ability of the classifier. The 12.8% gap between training and test accuracy of DKM is smaller than 16.4% of MSVM, so DKM has a better generalization ability (about 3.6%) than the common multi-kernel method MSVM. SVM, KFD, and MSVM are the optimal model obtained by grid search optimization algorithm, and DKM is the optimal model obtained by RACOS optimization algorithm. The running time of the optimal model is shown in Table 5, which shows that DKM is twice faster than MSVM. In conclusion, considering accuracy, generalization ability, and training time, DKM has the best comprehensive performance.

To further test the built DKM model, a blind well test is carried out using Well D13, in which there are five lithofacies, namely pebbled sandstone (PS), coarse sandstone (CS), medium sandstone (MS), fine sandstone (FS), and carbonate rock (CR). As shown in Fig. 10, the predicted lithofacies by DKM, KFD, MSVM, and SVM are: (1) 2682.1–2686.4 m, prediction of DKM is consistent with rock

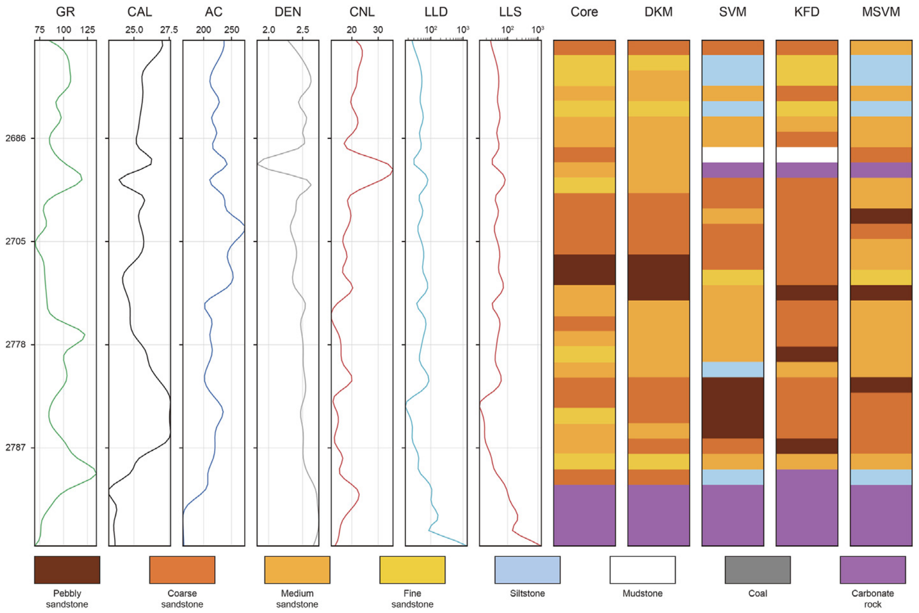


Fig. 10. Lithofacies identification of Well D13 by DKM, SVM, KFD and MSVM methods.

cores, SVM and MSVM misidentify fine sandstone (FS) as siltstone (S), KFD misidentifies medium sandstone (MS) as coarse sandstone (CS); (2) 2686.4–2704.8 m, prediction of DKM is basically consistent with rock cores, with only two misidentification errors, that is, fine sandstone (FS) and coarse sandstone (CS) are misidentified as medium sandstone (MS). Due to the complexity of lithology, SVM, KFD, and MSVM misidentified some sandstones (MS) and fine sandstone (FS) as other lithologies; (3) 2704.8–2778.2 m, the recognitions of DKM and MSVM are relatively good, SVM did not recognize pebbled sandstone (PS), and KFD incorrectly identified medium sandstone (MS) as pebbled sandstone (PS); (4) 2778.2–2787.3 m, all models failed to identify fine sandstone (FS), and SVM also incorrectly identified coarse sandstone (CS) as pebbled sandstone (PS); (5) 2787.3–2849.9 m, DKM predicted well, SVM, KFD, and MSVM incorrectly identified fine sandstone (FS) as medium sandstone (MS). Carbonate rock (CR) is well recognized by these methods. In general, DKM has a good capability of lithofacies identification, although some lithofacies identification is wrong. From the results in Fig. 10, fine sandstone is more difficult to identify than other lithologies and is easily misclassified as medium sandstone (MS) and siltstone (S). In general, most of the predicted lithofacies by DKM are consistent with core observation; DKM outperforms the other three methods (KFD, MSVM, and SVM); KFD and MSVM perform better than SVM.

Four quantitative metrics are used to evaluate the predictions of different methods on the blind well, which are accuracy, macro-precision, macro-recall, and macro-F1 score. (1) Accuracy is (TP + TN)/(TP + FP + TN + FN), which is an overall evaluation of all lithology categories and is the most commonly used performance indicator; (2) Precision is TP/(TP + FP), which measures the accuracy of each lithology predicted by the model. Macro-precision is the average of precisions corresponding to each lithology as positive category; (3) Recall is TP/(TP + FN), which measures the model's ability to find positive samples for each lithology. Similarly, macro-recall is the average of each recall corresponding to each lithology; (4) F1 score is (2 × precision × recall)/(precision + recall), which takes both precision and recall into account, and can reflect the accurate and complete capability of a prediction model. Macro-F1 score is also the average of F1 scores corresponding to different lithologies.

Table 6
Quantitative evaluation of different methods on a blind well test.

Model	Accuracy, %	Macro-precision, %	Macro-recall, %	Macro-F1 score, %
DKM	76	82	81	78
SVM	42	28	29	28
KFD	52	47	40	38
MSVM	39	29	32	29

As shown in Table 6, DKM is obviously superior to the other three kernel methods in this blind well test. Its accuracy, macro-precision, macro-recall, and macro-F1 score is 76%, 82%, 81%, and 78%, respectively. Compared with the other three kernel methods in the blind well test, there are improvements of about 24%, 35%, 41%, and 40% in accuracy, macro-precision, macro-recall, and macro-F1 score, respectively.

4. Discussions

The lithology label is important for the following lithofacies identification by machine learning methods. To build a stable and accurate model of lithofacies identification, more attention should be paid to quality control of labelled well log data. In this work, the lithology labels of well logs used are obtained based on core analysis from Huabei branch of Sinopec Group, which ensures the labeled data reliable.

4.1. Why is DKM suitable for identification problems of complex lithofacies

Lithofacies identification is usually a nonlinear classification problem, especially in the complex cases. Typically, multi-kernel methods can perform better than single-kernelled methods. This indicates that more powerful ability to deal with nonlinear features is helpful to distinguish lithofacies. Hence, deeply mining nonlinear features for distinguishing lithofacies is still needed to be further strengthened. DKM is a combination of kernel feature extraction and deep learning. From the aspect of theory, the addition of deep learning can enhance the ability of addressing nonlinear features of

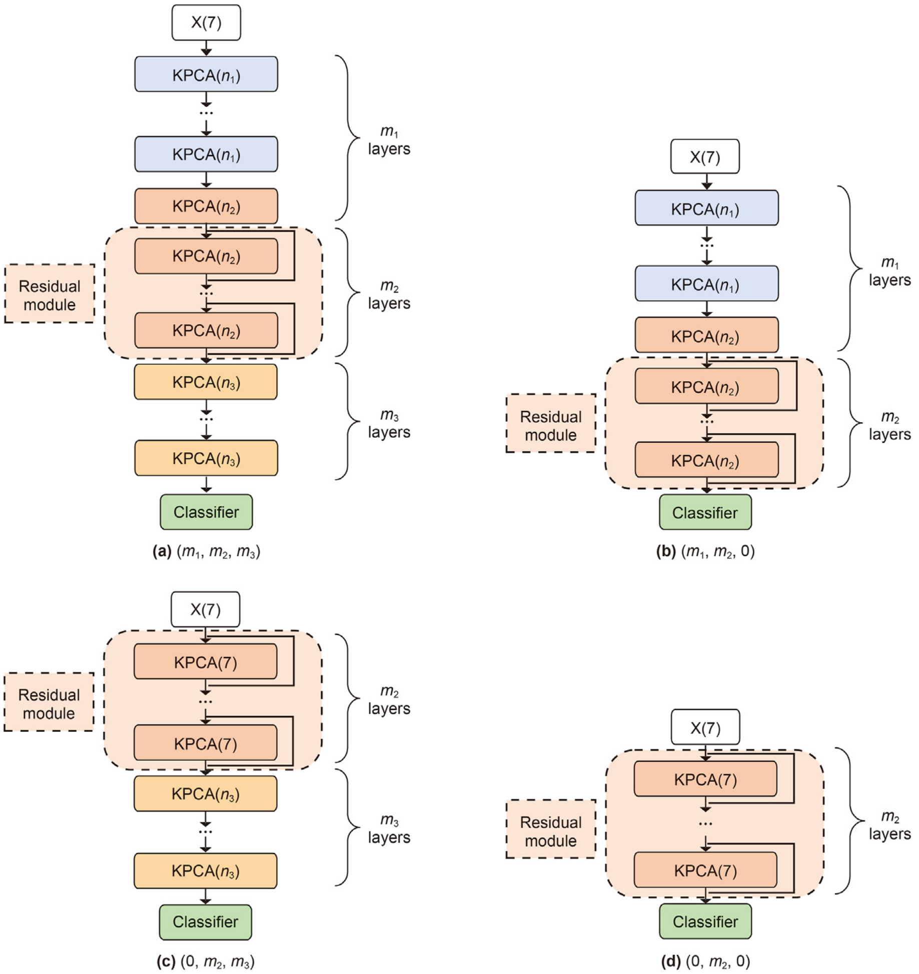


Fig. 11. Schematic diagram of model structures.

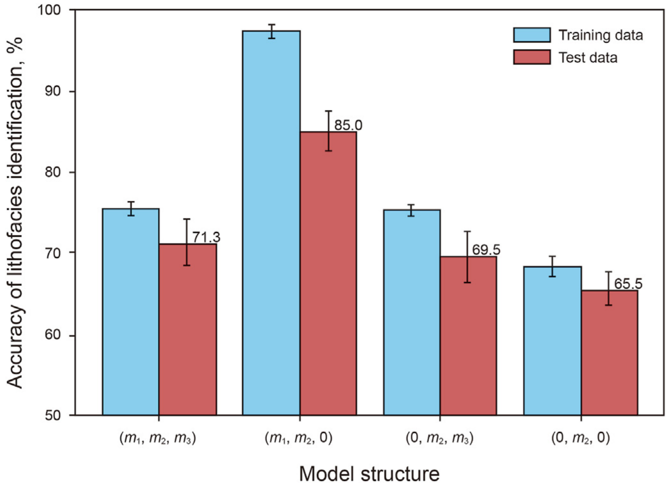


Fig. 12. Accuracy with error lines for different model structures.

lithofacies. This was proven in the lithofacies identification in this work, too. The outperformance of DKM compared with other common kernel methods indicates that DKM is suitable for identification problems of complex lithofacies.

In general, DKM has the following advantages: (1) Compared with deep neural network, DKM uses a gradient-free optimization method to automatically optimize parameters and structure, which can avoid complex tuning of parameters in models; (2) Residual units ensure DKM performs well in deep mining lithofacies features by reducing information loss using skip connections; (3) DKM inherits the advantages of kernel methods in nonlinear feature extraction; (4) DKM is an improvement of both neural networks and kernel methods, which let neural networks not limited to neurons, and kernel methods with the ability of deep mining in neural networks.

4.2. *Analyses on structures in DKM*

Different structures of DKM will have certain impacts on the prediction results, so it is very important to choose a suitable structure. In this work, four structures for DKM are considered as shown in Fig. 11. Fig. 11a shows the structure of (m_1, m_2, m_3), which is composed of $m_1 - 1$ KPCA layers with n_1 dimensions, the $m_1 -$ th KPCA layer with n_2 features, m_2 layers of KPCA residual units with dimension n_2, and m_3 KPCA layers with dimension n_3. Fig. 11b shows the structure of $(m_1, m_2, 0)$, which is composed of $m_1 - 1$ KPCA layers with n_1 dimensions, the $m_1 -$ th KPCA layer with n_2 features, and m_2 layers of KPCA residual units with dimension n_2. Fig. 11c shows the structure of $(0, m_2, m_3)$, which is composed of m_2

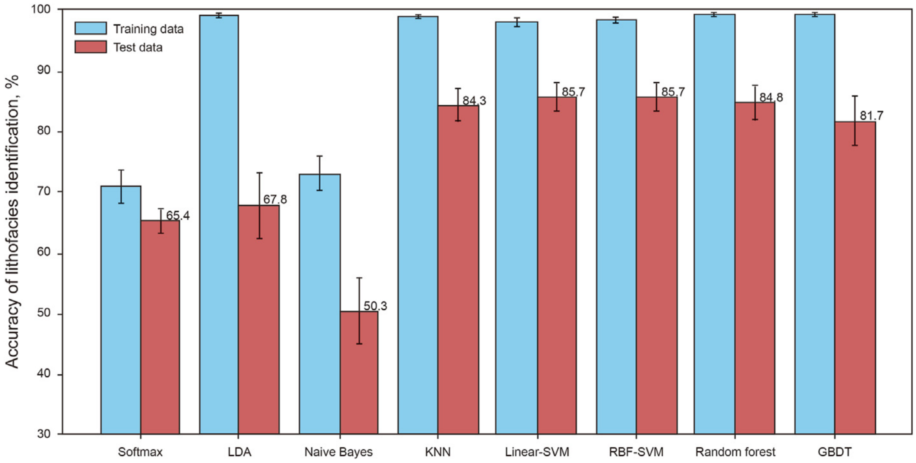


Fig. 13. Accuracy of lithofacies identification by different classifiers. The bar height and error line represent the mean accuracy and standard deviation of 20 repeats, respectively.

Table 7
Training time of different classifiers.

Classifier	Time
Softmax	4′6″
LDA	3′5″
Naive Bayes	5′39″
KNN	3′53″
Linear-SVM	4′5″
RBF-SVM	6′23″
Random forest	5′4″
GBDT	15′42″

layers of KPCA residual units with dimension 7, and m_3 KPCA layers with dimension n_3. Fig. 11d shows the structure of $(0, m_2, 0)$, which is composed of m_2 layers of KPCA residual units with dimension 7.

Theoretically, KPCA feature extraction and residual module can extract nonlinear features well, but different combinations have different effects on the model. Therefore, we transform and combine them into four different structures to explore which combination can better improve the model accuracy and which module has a significant impact on model prediction.

Each model structure was randomly trained 20 times. Fig. 12 shows the experimental results of different model structures. It can be seen that $(m_1, m_2, 0)$ is the optimal structure, on average, the accuracy of the test set is 85%. According to the model $(0, m_2, 0)$, it can be seen that the extracted information containing only residual structure is not sufficient, and the model has the problem of underfitting. Compared with model $(0, m_2, 0)$, model $(m_1, m_2, 0)$ added feature extraction module before residual module, and the average accuracy of test set was 19.5% higher, indicating that adding separate KPCA extraction module before residual structure and increasing feature dimension can improve feature extraction accuracy without losing original information. The average accuracy of model $(m_1, m_2, 0)$ and model (m_1, m_2, m_3) on the test set is 85% and 71.3%. It can be seen that after residual module extraction, if the dimensional-reduction KPCA module is added, effective information will be lost, resulting in reduced classification accuracy. The comparison between model (m_1, m_2, m_3) (71.4%) and model $(0, m_2, m_3)$ (69.5%) shows that the dimension enhancement before residual plays a key role in better feature extraction, while the residual plays a role of information supplement. Residual can effectively reduce information loss and improve the final prediction accuracy through information transfer across layers.

4.3. Analyses on classifiers in DKM

Based on Section 4.2, the optimal structure $(m_1, m_2, 0)$ is selected. In this Section, eight classifiers are selected to analyze the optimal classifiers for DKM in lithofacies identification. For each method, experiments using randomly selected training data are repeated 20 times. An optimal classifier for DKM should have high accuracy, good generalization ability, simplicity, and stability. The high accuracy means that the classifier has a high accuracy in test data. The good generalization ability means there is a small gap between accuracies of training and test data. The simplicity means the classifier has fewer parameters to determine and needs less time to predict. The stability means that the classifier has small standard deviation.

Accuracy of each classifier is shown in Fig. 13 while the consumed time is listed in Table 7.

As shown in Fig. 13, the accuracies of the five classifiers (KNN, 84.3%; linear-SVM, 85.7%; RBF-SVM, 85.7%; random forest, 84.8%; GBDT, 81.7%) are all more than 80% while those of Softmax (65.4%), LDA (67.8%) and Naive Bayes (50.3%) are all less than 70%. From the aspect of accuracy, Softmax, LDA and Naive Bayes will not be selected as the optimal classifier even though they are all simple and less time consuming.

In the five rest classifiers (KNN, linear-SVM, RBF-SVM, random forest, and GBDT), GBDT is most time consuming (15’42’’); the accuracy gap between training and test data is highest; the accuracy of GBDT is relatively low. Hence, GBDT will not be the optimal classifier either.

For the remaining four methods (KNN, linear-SVM, RBF-SVM, random forest), gaps between accuracies of training and test data are 15.0%, 12.5%, 12.8%, 14.4%, respectively. The generalization abilities of linear-SVM and RBF-SVM are better than those of KNN and random forest.

Accuracy standard deviations of KNN, linear-SVM, RBF-SVM and random forest are 2.6%, 2.4%, 2.3%, 2.8%, respectively. The stabilities of prediction of linear-SVM, RBF-SVM are a little better than those of KNN and random forest.

The consumed time of KNN, linear-SVM, RBF-SVM, and random forest is 3’53”, 4’5”, 6’23”, and 5’4”. They are between 3’ and 7’. The speed efficiencies of KNN, linear-SVM are a slightly better than RBF-SVM and random forest.

Except speed efficiency, RBF-SVM and linear-SVM perform similarly and best in accuracy of the eight classifiers. Noted that linear-SVM is simpler than RBF-SVM. According to Occam's Razor principle, the simpler the classifier selected, the better (Zahálka and

Železný, 2011). Hence, from this aspect, linear-SVM is more suitable for being the classifier in DKM than RBF-SVM. Because linear-SVM is also faster than RBF-SVM. Therefore, linear-SVM is finally selected as the classifier for DKM.

From the analyses abovementioned, a workflow to determine an optimal classifier can be summarized: (1) choose a series of candidate classifiers; (2) remove classifiers with low accuracy; (3) remove classifiers with high standard deviations of accuracy or bad generalization ability; (4) choose a simpler classifier with high speed as the optimal classifier for DKM. According to this workflow, an optimal classifier can be determined. It is not unique, which depends on the series of candidate classifiers.

Analyzing the advantages and disadvantages of different classifiers can give some hints to determining a good classifier for DKM. For example, linear-SVM has many advantages as a classifier in DKM since it's simple, fast, of good generalization ability, and so on. It can work well in most cases and it is recommended as classifier in DKM. However, if geologists want to improve DKM further, the workflow abovementioned to find an optimal classifier for DKM from a series of candidate classifiers by comparison experiments is needed.

4.4. Which are important parameters in DKM

Table 3 lists the parameters used by DKM, among which the most important parameters are n_2, m_2, and C, where n_2 represents the feature dimension of KPCA with residual structure, m_2 represents the number of KPCA layers with residual structure, C represents the penalty coefficient in the SVM model. A proper n_2 can map features from lower dimensions to higher dimensions in a linearly separable state, making it easier to classify categories, and in three dimensions it is shown that different categories can be divided by plane. m_2 controls the number of layers of extracted features with residual structure. A small number of layers makes extracted features insufficient, while a large number of layers makes extracted features more abstract. C is the penalty parameter in the Linear-SVM classifier. If the value of C is set large, SVM can better find the decision boundary of all training points, but it will make the training time of the model longer. If the set value of C is smaller, the SVM decision will be simpler, but the training accuracy will be reduced. C has a significant impact on the classifier, so it will affect the performance of DKM. In this paper, parameters n_2, m_2 and C are optimized to improve the performance of the DKM model and the accuracy of kernel methods for lithology identification.

4.5. Drawbacks and future work

It should be noted that DKM still has some issues to be addressed.

(1) In terms of parameter optimization, DKM may encounter difficulties in global optimization. Selecting an appropriate set of kernel parameters can improve model performance. We choose RACOS optimization algorithm, which can maintain a certain accuracy and has the advantage of high speed in high-dimensional feature space, but it may converge to a local minimum if the number of iterations is not enough, so it is not the best optimization method, and adopting a better optimization method in the future is the focus of research.

(2) In terms of the model feature extraction structure, we conducted comparative experiments and selected the optimal model structure. However, the KPCA feature extraction module of each layer uses a single kernel, so the model can try to use multiple kernels in the future to further improve the model generalization ability and improve the model performance.

(3) In terms of residual structure, only the model depth is considered, and the model width can be increased in the future to further improve the model performance.

5. Conclusions

In this paper, a deep kernel method (DKM) combining deep learning and a kernel method is proposed to improve lithofacies recognition. DKM adopts the gradient-free optimization method to automatically adjust parameters, which solves the problem of parameter tuning due to a large number of parameters and can quickly establish the lithofacies recognition model. To verify the effectiveness of this method in lithofacies identification, a series of comparative experiments were carried out using DKM and auxiliary kernel methods (KFD, SVM, MSVM). The optimal structure and classifier of DKM are also discussed. The following conclusions can be drawn:

(1) DKM (85.7%) outperforms SVM (77%), KFD (79.5%) and MSVM (82.8%) in accuracy. DKM has a better generalization ability (3.6%) than the commonly used multi-kernel method MSVM.

(2) DKM is about twice faster than that of MSVM, which has the best accuracy in the commonly used kernel methods (KFD, SVM, MSVM).

(3) $(m_1, m_2, 0)$ is the optimal model structure which is composed of $m_1 - 1$ KPCA layers with n_1 dimensions, the $m_1 -$ th KPCA layer with n_2 features, and m_2 layers of KPCA residual units with dimension n_2. After the input data, improving the feature dimension through KPCA and then adding residual module has a significant influence on improving the accuracy.

(4) Linear-SVM is the optimal classifier recommended for DKM in most cases. A workflow to determine an optimal classifier is proposed.

The future model can be further studied in kernel feature extraction structure and parameter optimization method to improve DKM.

Acknowledgement

This work was financially supported by the National Natural Science Foundation of China (Grant No. 42002134), China Postdoctoral Science Foundation (Grant No. 2021T140735) and Science Foundation of China University of Petroleum, Beijing (Grant Nos. 2462020XKJS02 and 2462020YXZZ004).

References

Al-Najjar, H.A.H., Pradhan, B., 2021. Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. Geosci. Front. 12 (2), 625–637. https://doi.org/10.1016/j.gsf.2020.09.002.

Anowar, F., Sadaoui, S., Selim, B., 2021. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). Comput. Sci. Rev. 40, 100378. https://doi.org/10.1016/j.cosrev.2021.100378.

Areiza-Laverde, H.J., Castro-Ospina, A.E., Hernández, M.L., et al., 2020. A novel method for objective selection of information sources using multi-kernel SVM and local scaling. Sensors 20 (14), 3919. https://doi.org/10.3390/s20143919.

Chen, X., Xu, S.Y., Li, S.M., et al., 2021. Identification of architectural elements based on SVM with PCA: a case study of sandy braided river reservoir in the Lamadian oilfield, Songliao Basin, NE China. J. Petrol. Sci. Eng. 198, 108247. https://doi.org/10.1016/j.petrol.2020.108247.

Cover, T.M., 1965. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Trans.

Electron.Comput. 14 (3), 326–334. https://doi.org/10.1109/PGEC.1965.264137. EC-.

Deng, C.X., Pan, H.P., Fang, S.N., et al., 2017. Support vector machine as an alternative method for lithology classification of crystalline rocks. J. Geophys. Eng. 14 (2), 341–349. https://doi.org/10.1088/1742-2140/aa5b5b.

Dong, S.Q., Wang, Z.Z., Zeng, L.B., 2016. Lithology identification using kernel Fisher discriminant analysis with well logs. J. Petrol. Sci. Eng. 143, 95–102. https://doi.org/10.1016/j.petrol.2016.02.017.

Dong, S.Q., Zeng, L.B., Che, X.H., et al., 2022a. Application of artificial intelligence in fracture identification using well logs in tight reservoirs. J. Earth Sci. 1–23. https://doi.org/10.3799/dqkx.2022.088 (in Chinese).

Dong, S.Q., Zeng, L.B., Du, X.Y., et al., 2022b. Lithofacies identification in carbonate reservoirs by multiple kernel Fisher discriminant analysis using conventional well logs: a case study in A oilfield, Zagros Basin, Iraq. J. Petrol. Sci. Eng. 210, 110081. https://doi.org/10.1016/j.petrol.2021.110081.

Dong, S.Q., Zeng, L.B., Liu, J.J., et al., 2020a. Fracture identification in tight reservoirs by multiple kernel Fisher discriminant analysis using conventional logs. Interpretation 8 (4), SP215–SP225. https://doi.org/10.1190/INT-2020-0048.1.

Dong, S.Q., Zeng, L.B., Lyu, W.Y., et al., 2020b. Fracture identification and evaluation using conventional logs in tight sandstones: a case study in the Ordos Basin, China. Energy Geosci. 1 (3–4), 115–123. https://doi.org/10.1016/j.engeos.2020.06.003.

Dong, S.Q., Zeng, L.B., Lyu, W.Y., et al., 2020c. Fracture identification by semi-supervised learning using conventional logs in tight sandstones of Ordos Basin, China. J. Nat. Gas Sci. Eng. 76, 103131. https://doi.org/10.1016/j.jngse.2019.103131.

Esmaeili, M., Ahmadi, M., Kazemi, A., 2020. Kernel-based two-dimensional principal component analysis applied for parameterization in history matching. J. Petrol. Sci. Eng. 191, 107134. https://doi.org/10.1016/j.petrol.2020.107134.

Farouk, S., Sen, S., Ganguli, S.S., et al., 2021. Petrophysical assessment and permeability modeling utilizing core data and machine learning approaches-A study from the Badr El Din-1 field, Egypt. Mar. Petrol. Geol. 133, 105265. https://doi.org/10.1016/j.marpetgeo.2021.105265.

Ge, X.M., Fan, Y.R., Zhu, X.J., et al., 2014. A method to differentiate degree of volcanic reservoir fracture development using conventional well logging data-An application of kernel principal component analysis (KPCA) and multifractal detrended fluctuation analysis (MFDFA). IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 7 (12), 4972–4978. https://doi.org/10.1109/JSTARS.2014.2319392.

Gu, Y.F., Liu, H., 2016. Sample-screening MKL method via boosting strategy for hyperspectral image classification. Neurocomputing 173, 1630–1639. https://doi.org/10.1016/j.neucom.2015.09.035.

He, H.S., Kong, F.Y., Tan, J.D., 2016. DietCam: multiview food recognition using a multi-kernel SVM. IEEE J. Biomed. Health Inform. 20 (3), 848–855. https://doi.org/10.1109/JBHI.2015.2419251.

Heidary, M., 2015. The use of kernel principal component analysis and discrete wavelet transform to determine the gas and oil interface. J. Geophys. Eng. 12 (3), 386–399. https://doi.org/10.1088/1742-2132/12/3/386.

Hou, E., Wen, Q., Ye, Z., et al., 2020. Height prediction of water-flowing fracture zone with a genetic-algorithm support-vector-machine method. Int. J. Coal. Sci. Technol. 7 (4), 740–751. https://doi.org/10.1007/s40789-020-00363-8.

Lan, X.X., Zou, C.C., Kang, Z.H., et al., 2021. Log facies identification in carbonate reservoirs using multiclass semi-supervised learning strategy. Fuel 302, 121145. https://doi.org/10.1016/j.fuel.2021.121145.

Lanckriet, G.R.G., Cristianini, N., Bartlett, P., et al., 2004. Learning the kernel matrix with semidefinite programming. J. Mach. Learn. Res. 5 (1), 27–72.

Li, C.S., Tang, G., Xue, X.M., et al., 2020. The short-term interval prediction of wind power using the deep learning model with gradient descend optimization. Renew. Energy 155, 197–211. https://doi.org/10.1016/j.renene.2020.03.098.

Li, Q.Z., Fu, Q., Zou, Y., et al., 2021. Evaluation of livable city based on GIS and PSO-SVM: a case study of Hunan Province. Int. J. Pattern Recogn. Artif. Intell. 35, 2159030. https://doi.org/10.1142/S0218001421590308, 08.

Liang, W.Z., Luo, S.Z., Zhao, G.Y., et al., 2020. Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. Mathematics 8 (5), 765. https://doi.org/10.3390/math8050765.

Lin, F., Wang, J.B., Zhang, N., et al., 2017. Multi-kernel learning for multivariate performance measures optimization. Neural Comput. Appl. 28 (8), 2075–2087. https://doi.org/10.1007/s00521-015-2164-9.

Liu, J.J., Wu, C.Z., Wu, G.N., et al., 2015. A novel differential search algorithm and applications for structure design. Appl. Math. Comput. 268, 246–269. https://doi.org/10.1016/j.amc.2015.06.036.

Liu, X.Y., Chen, X.H., Li, J.Y., et al., 2020a. Facies identification based on multikernel relevance vector machine. IEEE Trans. Geosci. Rem. Sens. 58 (10), 7269–7282. https://doi.org/10.1109/TGRS.2020.2981687.

Liu, X.Y., Zhou, L., Chen, X.H., et al., 2020b. Lithofacies identification using support vector machine based on local deep multi-kernel learning. Petrol. Sci. 17 (4), 954–966. https://doi.org/10.1007/s12182-020-00474-6.

Liu, Y., Ji, Y., Liu, D., et al., 2021. A new method for runoff prediction error correction based on LS-SVM and a 4D copula joint distribution. J. Hydrol. 598, 126223. https://doi.org/10.1016/j.jhydrol.2021.126223.

Malik, H., Mishra, S., 2016. Proximal support vector machine (PSVM) based imbalance fault diagnosis of wind turbine using generator current signals. Energy Proc. 90, 593–603. https://doi.org/10.1016/j.egypro.2016.11.228.

Mansouri-Daneshvar, P., Moussavi-Harami, R., Mahboubi, A., et al., 2015. Sequence stratigraphy of the petroliferous dariyan formation (aptian) in qeshm island and offshore (southern Iran). Petrol. Sci. 12 (2), 232–251. https://doi.org/10.1007/s12182-015-0027-8.

Mercer, J., 1909. Functions of positive and negative type, and their connection with the theory of integral equations. Proc. Roy. Soc. Lond. 209, 415–446. https://doi.org/10.1098/rsta.1909.0016.

Mohebian, R., Riahi, M.A., Afjeh, M., 2018. Detection of the gas-bearing zone in a carbonate reservoir using multi-class relevance vector machines (RVM): comparison of its performance with SVM and PNN. Carbonates Evaporites 33 (3), 347–357. https://doi.org/10.1007/s13146-017-0411-0.

Ortiz-García, E.G., Salcedo-Sanz, S., Pérez-Bellido, Á.M., et al., 2009. Improving the training time of support vector regression algorithms through novel hyper-parameters search space reductions. Neurocomputing 72 (16–18), 3683–3691. https://doi.org/10.1016/j.neucom.2009.07.009.

Pablo, D.G., Miguel, L., Jaume, B., et al., 2016. GPU-SME-kNN: scalable and memory efficient kNN and lazy learning using GPUs. Inf. Sci. 373, 165–182. https://doi.org/10.1016/j.ins.2016.08.089.

Pan, B.Z., Wang, X.R., Guo, Y.H., et al., 2022. Study on reservoir characteristics and evaluation methods of altered igneous reservoirs in Songliao Basin, China. J. Petrol. Sci. Eng. 212, 110266. https://doi.org/10.1016/j.petrol.2022.110266.

Phillips, T., Abdulla, W., 2021. Developing a new ensemble approach with multi-class SVMs for Manuka honey quality classification. Appl. Soft Comput. 111, 107710. https://doi.org/10.1016/j.asoc.2021.107710.

Qiao, X., Chang, F., 2021. Underground location algorithm based on random forest and environmental factor compensation. Int. J. Coal. Sci. Technol. 8 (5), 1108–1117. https://doi.org/10.1007/s40789-021-00418-4.

Qu, F.Z., Jiang, Q., Jin, G.Z., et al., 2020. Mud pulse signal demodulation based on support vector machines and particle swarm optimization. J. Petrol. Sci. Eng. 193, 107432. https://doi.org/10.1016/j.petrol.2020.107432.

Schölkopf, B., Smola, A., Müller, K., 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. 10 (5), 1299–1319. https://doi.org/10.1162/089976698300017467.

Shen, C.B., Asante-Okyere, S., Yevenyo Ziggah, Y., et al., 2019. Group method of data handling (GMDH) lithology identification based on wavelet analysis and dimensionality reduction as well log data pre-processing techniques. Energies 12 (8), 1509. https://doi.org/10.3390/en12081509.

Shi, J.X., Zeng, L.B., Dong, S.Q., et al., 2020. Identification of coal structures using geophysical logging data in Qinshui Basin, China: investigation by kernel Fisher discriminant analysis. Int. J. Coal Geol. 217, 103314. https://doi.org/10.1016/j.coal.2019.103314.

Sun, Z.X., Jiang, B.S., Li, X.L., et al., 2020. A data-driven approach for lithology identification based on parameter-optimized ensemble learning. Energies 13 (15), 3903. https://doi.org/10.3390/en13153903.

Suykens, J.A.K., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural Process. Lett. 9 (3), 293–300. https://doi.org/10.1023/A:1018628609742.

Tambe, S.S., Naniwadekar, M., Tiwary, S., et al., 2018. Prediction of coal ash fusion temperatures using computational intelligence based models. Int. J. Coal. Sci. Technol. 5 (4), 486–507. https://doi.org/10.1007/s40789-018-0213-6.

Tang, T.L., Chen, S.Y., Zhao, M., et al., 2019. Very large-scale data classification based on K-means clustering and multi-kernel SVM. Soft Comput. 23 (11), 3793–3801. https://doi.org/10.1007/s00500-018-3041-0.

Tharwat, A., 2019. Parameter investigation of support vector machine classifier with kernel functions. Knowl. Inf. Syst. 61 (3), 1269–1302. https://doi.org/10.1007/s10115-019-01335-4.

Tharwat, A., Hassanien, A.E., 2019. Quantum-behaved particle swarm optimization for parameter optimization of support vector machine. J. Classif. 36 (3), 576–598. https://doi.org/10.1007/s00357-018-9299-1.

Tipping, M.E., 2001. Sparse bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244. https://doi.org/10.1162/15324430152748236.

Wang, H., Peng, M.J., Yu, Y., et al., 2021. Fault identification and diagnosis based on KPCA and similarity clustering for nuclear power plants. Ann. Nucl. Energy 150, 107786. https://doi.org/10.1016/j.anucene.2020.107786.

Wang, Q.S., Zhang, X.J., Tang, B., et al., 2021. Lithology identification technology using BP neural network based on XRF. Acta Geophys. 69 (6), 2231–2240. https://doi.org/10.1007/s11600-021-00665-8.

Wang, T.H., Zhang, L., Hu, W.Y., 2021. Bridging deep and multiple kernel learning: a review. Inf. Fusion 67, 3–13. https://doi.org/10.1016/j.inffus.2020.10.002.

Wang, X.X., Zhang, F., Li, S.H., et al., 2021. The architectural surfaces characteristics of sandy braided river reservoirs, case study in Gudong oil field. China. Geofluids. 2021, 1–12. https://doi.org/10.1155/2021/8821711.

Wu, W., Jing, X.Y., Du, W.C., et al., 2021. Learning dynamics of gradient descent optimization in deep neural networks. Sci. China Inf. Sci. 64 (5), 150102. https://doi.org/10.1007/s11432-020-3163-0.

Xiao, Z.H., Jiang, W., Sun, B., et al., 2020. Quantitative identification of coal texture using the support vector machine with geophysical logging data: a case study using medium-rank coal from the Panjiang, Guizhou, China. Interpretation 8 (4), T753–T762. https://doi.org/10.1190/INT-2019-0237.1.

Xie, Y.X., Zhu, C.Y., Zhou, W., et al., 2018. Evaluation of machine learning methods for formation lithology identification: a comparison of tuning processes and model performances. J. Petrol. Sci. Eng. 160, 182–193. https://doi.org/10.1016/j.petrol.2017.10.028.

Xu, K., Sun, Z.D., 2018. Seismic interpolation based on a random forest method, 01 Petrol. Sci. Bull. 3, 22–31. https://doi.org/10.3969/j.issn.2096-1693.2018.01.003 (in Chinese).

Xu, L., Hou, L., Zhu, Z.Y., et al., 2021. Prediction of oilfield produced water treatment based on a two-layer decomposition technique and modified SVM, 03 Petrol.

Sci. Bull. 6, 505–515. https://doi.org/10.3969/j.issn.2096-1693.2021.03.041 (in Chinese).

Yan, F., Kittler, J., Mikolajczyk, K., et al., 2012. Non-sparse multiple kernel Fisher discriminant analysis. J. Mach. Learn. Res. 13 (21), 607–642.

Yates, D., Islam, M.Z., 2021. FastForest: increasing random forest processing speed while maintaining accuracy. Inf. Sci. 557, 130–152. https://doi.org/10.1016/j.ins.2020.12.067.

Yin, S., Yin, J.P., 2016. Tuning kernel parameters for SVM based on expected square distance ratio. Inf. Sci. 370–371, 92–102. https://doi.org/10.1016/j.ins.2016.07.047.

Yu, H.Y., Xu, Z., Wang, Y.L., et al., 2021. The use of KPCA over subspaces for cross-scale superpixel based hyperspectral image classification. Remote Sens. Lett. 12 (5), 470–477. https://doi.org/10.1080/2150704X.2021.1897180.

Yu, Y., Qian, H., Hu, Y.Q., 2016. Derivative-free optimization via classification. AAAI 16, 2286–2292.

Yuan, J.J., Wang, C.D., Zhou, Z.H., 2019. Study on refined control and prediction model of district heating station based on support vector machine. Energy 189, 116193. https://doi.org/10.1016/j.energy.2019.116193.

Zahálka, J., Železný, F., 2011. An experimental test of Occam's razor in classification. Mach. Learn. 82 (3), 475–481. https://doi.org/10.1007/s10994-010-5227-2.

Zhang, C.L., He, Y.G., Yuan, L.F., et al., 2017. Capacity prognostics of lithium-ion batteries using EMD denoising and multiple kernel RVM. IEEE Access 5, 12061–12070. https://doi.org/10.1109/ACCESS.2017.2716353.

Zhang, G.Y., Wang, Z.Z., Mohaghegh, S., et al., 2021. Pattern visualization and understanding of machine learning models for permeability prediction in tight sandstone reservoirs. J. Petrol. Sci. Eng. 200, 108142. https://doi.org/10.1016/j.petrol.2020.108142.

Zhang, H., Jiang, L.X., Yu, L.J., 2020. Class-specific attribute value weighting for Naive Bayes. Inf. Sci. 508, 260–274. https://doi.org/10.1016/j.ins.2019.08.071.

Zhao, T., Jayaram, V., Marfurt, K.J., et al., 2014. Lithofacies classification in Barnett Shale using proximal support vector machines. In: 2014 SEG Annual Meeting, pp. 1491–1495. https://doi.org/10.1190/segam2014-1210.1.

Zhou, J., Lin, H.F., Jin, H.W., et al., 2022. Cooperative prediction method of gas emission from mining face based on feature selection and machine learning. Int. J. Coal. Sci. Technol. 9 (1). https://doi.org/10.1007/s40789-022-00519-8.

Zhu, L.B., Chen, D., Feng, P.F., 2021. Equipment operational reliability evaluation method based on RVM and PCA-fused features. Math. Probl Eng. 2021, 1–9. https://doi.org/10.1155/2021/6687248.