

Received December 17, 2020, accepted December 18, 2020, date of publication December 30, 2020, date of current version March 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048078

# A Novel Intelligent Decision-Making Method of Shearer Drum Height Regulating Based on Neighborhood Rough Reduction and Selective Ensemble Learning

ZHENGXIONG LU<sup>1</sup>, WEI GUO, CHUANWEI ZHANG, SHUANFENG ZHAO, YUAN WANG, WUGANG ZHANG, MANZHI YANG, AND SHUAITIAN LI

School of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China

Corresponding author: Shuanfeng Zhao (zsf@xust.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0804310 and Grant 2018YFC0808 203, and in part by the National Natural Science Foundation of China under Grant 51805428.

**ABSTRACT** An intelligent shearer drum height regulating method is the key technology for mining at an unmanned coalface. In this study, a novel intelligent decision-making method of shearer drum height regulating is proposed, which makes a decision by selective ensemble the Kernel Extreme Learning Machine (KELM) with a self-learning ability. In this approach, the shearing coal process of the shearer is characterized based on the extended finite state machine. Transfer attributes are introduced to establish the decision information system of shearer drum height regulating. Then, propose a neighborhood rough reduction method is proposed to generate distinctive attribute subsets, which is applied to train the base classifiers based on the online KELM. Finally, we introduce an accuracy-guided forward search and post-pruning strategy to select part of the base classifiers for constructing an efficient and effective ensemble system of the shearer drum lifting prediction. For evaluating the proposed method, four evaluation metrics are used: accuracy, precision, recall rate and the F1-score, which are the most popular metrics for evaluating the performance of a classifier. We use the ten-fold cross validation method to optimize the hyperparameters. The proposed method is compared in two different scenarios: 1) three different classes of base classifier algorithms which including the Support Vector Machines (SVM), Support Vector Machines (CART) and K-NearestNeighbor (KNN) are used, and 2) two traditional ensemble methods including the bagging and random subspace. The proposed method is performed on the field datasets and the experimental results reveal that the method is effective in comparison to other approaches for shearer drum lifting prediction.

**INDEX TERMS** Shearer, machine learning, classifier selective ensemble, neighborhood rough set, KELM, drum lifting prediction.

## I. INTRODUCTION

### A. BACKGROUND

As the coal mining depth of underground increases, the disasters such as gas explosion, rock collapse and water inrush occur frequently in the process of the coal mining, which seriously threaten the lives of coal miners in full mechanized coalface [1]. To improve this situation, the unmanned fully mechanized coal mining method is considered as an effective

The associate editor coordinating the review of this manuscript and approving it for publication was Gongbo Zhou<sup>1</sup>.

way to solve this problem for its high recovery ratio and extremely low mortality rate in recent years [2], [3]. The intelligent shearer, one of the main equipment in fully mechanized coalface, is a key equipment when using this method. However, it has always been difficult for the shearer drum height intelligent regulating to realize the intelligence of the shearer, since the natural occurrence boundary of coal seams is very irregular due to sinking roof rocks and rising floor rocks frequently.

This poses a challenge to make the shearer drum height adapt to the undulating changes of the coal roof.

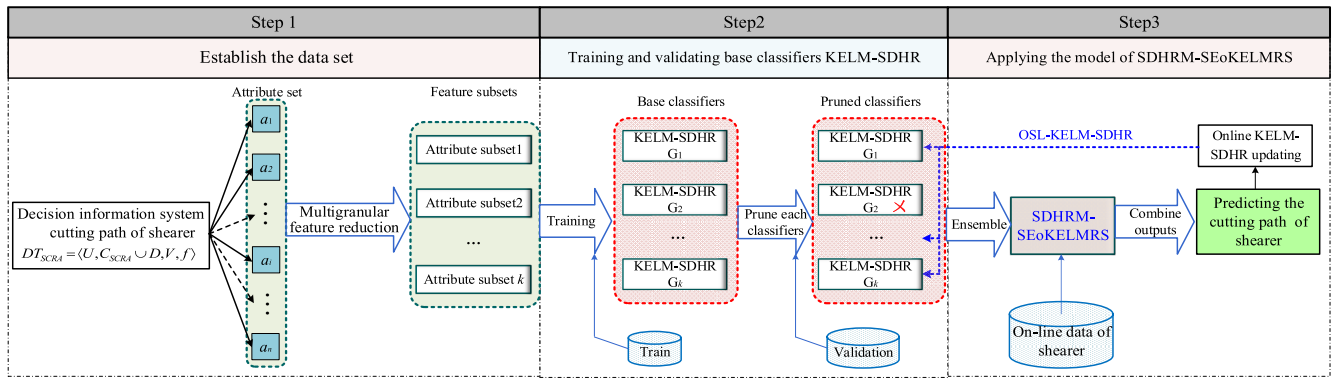


FIGURE 1. Steps of the proposed method.

A number of techniques for the shearer drum height intelligent regulating have been devised so far [4], [5]. Roughly speaking, we can divide these techniques into two kinds: simple feedback regulating based on identification of the coal-rock interface and intelligent regulating based on multi-information fusion. The simple feedback regulating refers to making use of the information of the coal-rock interface from the sensors to achieve drum height regulating directly. Widely researched simple feedback control techniques include the control methods based on the gamma ray instrument [6], the radar coal thickness sensor [7], the thermal infrared-based coal seam tracking [8]. However, these technologies, which are based on simple drum height feedback, have not been widely applied due to the structural complexity of the coal seam, and technical problems related to identification of the coal-rock interface. Furthermore, stable and practical online identification of coal-rock interface is a complex problem, which will influence the simple and direct feedback regulating of the shearer for a long time.

To address these problems, the shearer drum height automatic regulating method based on multi-sensor information fusion can be considered a reasonable alternative. With the development of computational technologies, the shearer drum height intelligent regulating methods based on field data started to attract more attention of researchers. Chen *et al.* [9] developed a prediction method for the shearing trajectory using a LSTM neural network, in which the historical data of shearer drum height was used as the only input information. Li *et al.* [10] presented an automatic regulating method based on the grey-Markov model. The predicted height data of shearer drum were used to build the state transfer probability matrix of the Markov chain to achieve better precision and stability. However, Chen and Li did not achieve the expected results because the single input information hardly described the complex coal seam occurrence. Fan *et al.* [11] proposed a shearer intelligent height adjusting system control method based on Dynamic Fuzzy Neural Networks (D-FNN). However, the input parameters of D-FNN obtain from a mathematical model, which can be hardly adaptable to all cases. Si *et al.* [12] combined a CNN and Dempster-Shafer

evidence theory to propose an intelligent multisensor data fusion and recognition method. By analyzing the vibration acceleration and current signals of six shearing conditions, the fusion algorithm predicted the shearing trajectory of the shearer. Wang *et al.* [13] obtained a model of self-adaptive adjustment height of the shearer drum based on artificial immunity and memory cutting. In this work, the shearing and working parameters of the shearer were considered to adapt the geological conditions of the coal-seam boundary changes. Yang and Xiong [14] also proposed an adaptive regulating height model of shearer drum, with two wavelet neural networks as identifier and controller respectively. Application results show that this control system is more effective than those based on normal neural network. Wang and Zhang [15] established an intelligent height adjustment model for shearer drums based on multi-sensor information fusion derived from the principle of minimum fuzzy entropy by testing vibration, current, acoustic emission and infrared signals during coal shearing. Xu *et al.* [16] proposed a self-adaptive cutting strategy using fuzzy theory, which can automatically adjust drum height as well as judge whether the shearer is cutting rocks. However, it is a challenge to define an appropriate fuzzy set and associated membership functions in such fuzzy logic-based models. However, none of these methods satisfies the expected standards of high accuracy because the existing methods have some disadvantages. First, the attributes describing changes of the coal seam from a local perspective are not considered. Second, there are methods depend on a single decision model which is difficult to effectively guide the shearer to intelligently adjust drum height in a fully mechanized coalface under complex and changeable geological conditions.

### B. RESEARCH MOTIVATIONS

The research motivation of this study is twofold: employment of the migration features and neighborhood rough reduction for describing and extracting the change law of the coal-rock interface, and the utilization of the selective ensemble learning methods-based the online self-optimizing KELM in height prediction of the shearer drum. The

regulating pattern of the shearer drum height is classified into three types, that is, increasing its height, decreasing its height and holding its height. From this perspective, this research transforms the problem of the shearer drum height prediction into the problem of pattern recognition (multiclass classification). Generally, an intelligent shearer height regulating classification system framework based on multisensor information fusion can be divided into three stages namely (1) data acquisition, defining the perceptual parameters and data preprocessing, (2) classification and (3) controller. The data is collected from sensing system of the operating shearer. Now, the perceptual parameters which used to be the shearer height regulating come from the shearing and working parameters of the shearer [9]–[11], [12]. However, these parameters cannot describe the change law of the coal-rock interface to the cutting height adjusting system of the shearer drum. Therefore, we propose the migration features of the shearer working state to construct the complete parameters set, which describe the change law of the coal-rock interface from the local perspective. It is worth noting that there may be redundant parameters in the parameters set of field data because of using the migration features, which would constrain the efficiency and accuracy of the shearer height regulating model. In order to handle these problems and increase the classification accuracy, the appropriate parameters (attributes) reduction techniques must be used. Rough set theory, which was introduced by Pawlak [17], has attracted much attention. This methodology proves to be a powerful tool for attribute reduction [18], [19]. As we know, most of the applications select the reduction with the fewest attributes to construct a classifier at present [20], [21]. However, the information hidden in other reducts is wasted in this case. In order to avoid the decrease of accuracy of the height prediction caused by information loss, an essential technique needs to be employed.

Furthermore, a suitable classifier with a better performance must be employed to predict the regulating pattern of the height of the shearer drum. Several studies on developing ensemble models employing single machine learning methods are found in literature [22], [23]. Ensemble machine learning techniques are methods which combine opinions of the multiple learners to achieve a better performance. It allows using a group of simple predictors while achieving a better classification performance. Tsai [24] stated that the ensemble models based on the combination of diverse classifiers accomplish a better classification performance by eliminating the other classifiers' errors. Vijaya and Sivasankar [25] proposed a signal classification model to combine feature selection with the bagging ensemble classifiers to enhance performance in terms of training time and accuracy. Hu *et al.* [26] presented an ensemble classifier by classifiers built on the different attribute reduction subset by sampling randomly from the original feature set, which selected part of the base classifiers. Saha *et al.* [27] used SVM, Random Forest, Decision Tree and NB to build an ensemble learning method based on majority voting for

prediction of protein interactions. In multiclass classification, Zhang *et al.* [28] used different classifier models to build an ensemble learning method for prediction of rockburst intensity and the algorithm using an ensemble classifier revealed a better classification performance. Hence, the main motivation of this study is to employ the ensemble classifiers in the shearer drum intelligent height adjusting to achieve a better prediction performance.

In this paper, a novel intelligent decision-making method of shearer drum height regulating is proposed based on neighborhood rough reduction and selective ensemble learning. The model can utilize the useful attributes information more effectively and it improves the efficiency and accuracy of the height prediction with KELM fast learning. Furthermore, enhancement of the fast learning and generalization performance of the proposed model is also achieved by the online self-learning KELM, which not only uses the general classification algorithms. The main contributions of this paper are as follows:

(1) First, a novel intelligent decision-making method of shearer drum height regulating is proposed based on neighborhood rough reduction and selective ensemble learning.

(2) Second, we propose the migration features to describe the change law of the coal-rock interface from the local perspective to improve the generalization performance of model with irregular changes of coal-rock interface.

(3) Third, we selectively combine partial reductions that redundant attributes are removed using the attribute reduction based on multi-granularity neighborhood rough set, where the information hidden in reductions is utilized effectively for improving the accuracy of shearer height regulating.

(4) Finally, the online self-optimizing KELM algorithm is used in the ensemble classifier to achieve self-adaptive adjustment of the shearer drum height regulating model.

Besides, various experiments are conducted in order to compare the proposed method and the Bagging and Random Subset ensemble classifiers performance for the shearer drum height regulating.

The rest of this paper is organized as follows. Section 2 introduces the modeling process of the intelligent decision-making model of shearer drum height regulating. Section 3 describes the full experiments including performance evaluation, hyperparameters tuning, and verifying the performance of the proposed method. The main results of this study are summarized and discussed in Section 4.

## II. METHODOLOGY

Fig 1 depicts the proposed method. The main ideas of the shearer drum height regulating method based on the selective ensemble of online KELM-based rough subspaces (SDHRM-SEoKELMRS) are as follows. (1) Field data preprocessing. First, the parameters (attributes) set of the cutting and working state in the shearer process, including general attributes and migration attributes, is constructed. The decision information system for the shearer drum height regulating is established based on field data. Second, the

different attribute subsets are obtained by the attribute reduction technology based on the multi-granular neighborhood rough set. Finally, the training sample sets and test sample sets are established corresponding to each attribute subspace. (2) Training and validating the base classifiers based on the kernel extreme learning machine. And then, an accuracy-guided forward search and post-pruning strategy is used to select parts of base classifiers for ensemble systems. (3) Application and online self-optimization of the SDHRM-SEoKELMRS model. Different modules are described in the following subsections.

**A. ESTABLISHMENT OF THE SDHR ATTRIBUTE SPACE**

**1) REFINED REPRESENTATION OF THE COAL SHEARING PROCESS OF THE SHEARER**

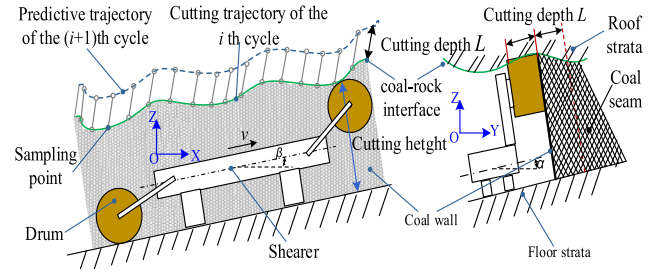
The shearer is one of the key equipment in fully mechanized coalface, whose main tasks are shearing and loading coal. The shearer adjusts the drum to adapt to change of coal seam thickness in fully mechanized coalface through the automatic height adjustment device, so that it can cut the coal seam along the coal-rock interface, and its principles are shown in Fig 2. X is the running direction of the shearer in the working face. Y is the advancing direction of the shearer in the working face. Z is the cutting height direction of the shearer.

Digital representation of the shearer shearing process is the basis for intelligent height adjustment of the shearer drum based on data-driven. Mataric [29] found that the interaction between an agent and the working environment can be modeled as synchronous finite-state automata. Thus, we propose a digital method for describing the cutting coal process of shearer based on an extended finite-state machine. In this way, we establish the mapping relationship between the state parameters and the behaviors of the shearer drum.

*Definition 1:* The extended finite-state machine of the cutting coal process of the shearer, formulated as a six-tuple  $SCEFSM = (Q', \Sigma', D, W, \delta, \lambda)$ , where  $Q' = \{q'_1, \dots, q'_n\}$  is the finite sensing parameters set of the cutting coal process of the shearer,  $\Sigma' = \{\sigma'_1, \sigma'_2, \dots, \sigma'_n\}$  is the value set of  $Q'$ ,  $D$  is the set of behavior parameters of the shearer drum,  $W$  is the value range of the behavior parameters,  $\delta = Q' \times \Sigma' \rightarrow Q'$  is the state transition function of the shearer, and  $\lambda = Q' \times \Sigma' \rightarrow W$  is the mapping between the finite sensing parameters and the behavior parameters. In addition, a three-tuple  $S = \{Q', \Delta, D\}$  is defined as the pattern, where  $\Delta = \{\Delta_1, \Delta_2, \dots, \Delta_n\}$  is a set of the migration features of the cutting coal process of the shearer.  $C = \{Q', \Delta\}$  is defined as the set of condition attributes of the decision information system for the intelligent height regulating of the shearer.

**2) CONSTRUCTION OF THE ATTRIBUTES SET**

It is important to realize intelligent decision-making of the shearer drum height regulating based on multi-information fusion to establish the mapping relationship between the state parameters, the change law of the coal-rock interface and the behaviors of the shearer drum. We propose the transfer



**FIGURE 2. Work cycle of the shearer.**

attribute set  $\Delta$  to construct the attribute set, which can make the pattern space  $S$  of the decision information system of the shearer drum height regulating.

The attribute set of the decision information system of the shearer drum height regulating needs to satisfy the essential conditions as follows: (1) the scale of the attributes does not change as the system changes, (2) there is a high correlation between the attributes and the behavior modes of the shearer drum, and (3) the attributes should be as complete as possible to reflect the changes of working state of the shearer drum. We establish the attributes set which is shown in Table 1 based on these three principles and engineering applications.

It can be seen that the attributes are not affected by the system scale in Table 1. Furthermore, the attributes are mainly the physical parameters reflecting the changes of working state of the shearer drum. The parameters  $c_1$ - $c_5$  reveal the information of position and attitude of the shearer. The parameters  $c_6$ - $c_{14}$  reflect the working information of the shearer. We define the change of the parameters  $c_6$ - $c_{14}$  as the migration features. The parameters  $c_{18}$ - $c_{26}$  are the migration attributes of the operational process, which reveals the evolution law of the coal-rock interface from the

local perspective when the cutting trajectory of shearer drum changes in the time and space. The values of the migration features are obtained by the difference in the value of the parameters  $c_6$ - $c_{14}$  between adjacent sampling periods. Based on the attribute set in Table 1, in order to analyze easily, we establish the decision information system of the shearer drum height regulating mode.

*Definition 2:* The decision information system of the shearer drum height regulating, DIS-SDHR is formulated as a five-tuple  $DIS_{SDHR} = (U, A, V, f)$ , where  $U = \{x_1, x_2, \dots, x_n\}$  is a nonempty set including finite objects  $x_i$ ,  $i = 1, 2, \dots, n$ .  $A = C \cup \{d\} \neq \emptyset$  and  $C \cap \{d\} = \emptyset$  are the set of attributes of the objects,  $C = \{c_1, c_2, \dots, c_{26}\}$  is the set of condition attributes, and  $d$  is the decision attribute, which indicates the shearer drum adjustment pattern in every sample. When the height of the drum increases, the value of  $d$  is 1. In contrast, the value of  $d$  is 2 when the height of the drum decreases. The value of  $d$  is 3 when the drum height is constant. Where  $V = \{v_1, v_2, \dots, v_{26}, v_d\}$  is the value range of the attributes,  $f$  is the information function, and  $f : U \times A \rightarrow V$  is used to define the attribute values of the sample.

TABLE 1. Coding attributes in the process shearing coal of shearer.

Condition attributes	Parameter name and unit	Condition attributes	Parameter name and unit
$c_1$	Shearer position /m	$c_{14}$	Current of right shearing motor /A
$c_2$	Pitch angle /°	$c_{15}$	Pitch angle variation /°
$c_3$	Roll angle /°	$c_{16}$	Roll angle variation /°
$c_4$	Height of left drum /m	$c_{17}$	variation of height of left drum /m
$c_5$	Height of right drum /m	$c_{18}$	variation of height of right drum /m
$c_6$	Pulling speed /m·min <sup>-1</sup>	$c_{19}$	Pulling speed variation /m·min <sup>-1</sup>
$c_7$	Traction direction	$c_{20}$	Current variation of left traction motor /A
$c_8$	Current of left traction motor /A	$c_{21}$	Temperature variation of left traction motor /°C
$c_9$	Temperature of left traction motor /°C	$c_{22}$	Right traction motor current variation/A
$c_{10}$	Right traction motor current /A	$c_{23}$	Temperature variation of right traction motor /°C
$c_{11}$	Temperature of right traction motor /°C	$c_{24}$	Current variation of left shearing motor /A
$c_{12}$	Current of left shearing motor /A	$c_{25}$	Temperature variation of left shearing motor /°C
$c_{13}$	Temperature of left shearing motor /°C	$c_{26}$	Current variation of right shearing motor /A

**B. CONSTRUCTION OF THE ROUGH SUBSETS**

The key problem using the attribute subsets ensemble method to adjust height of the shearer drum is how to get a set of attribute subset with good predicting power. In this section, we employ the multigranular neighborhood rough sets to get the attribute subsets based on the decision information system of the shearer drum height regulating.

**1) MULTIGRANULAR ATTRIBUTE REDUCTION FOR DIS-SDHR**

The rough set theory was first proposed by Prof. Pawlak for processing imprecision, vagueness, and uncertainty of data [17]. Attribute reduction is an important application of rough set theory that is used to remove the redundant and irrelevant attributes in attribute set. Attribute values of the decision information system of the shearer drum height regulating are continuous and discrete data. However, Pawlaw rough sets can only deal with discrete data. In order to deal with this problem, the neighborhood rough set and attribute reduction technology [30] are used to establish the subspaces of attribute set by changing the neighborhood radius.

Given a decision information system of the shearer drum height regulating  $DIS_{SDHR} = \langle U, C \cup D, V, f \rangle, B \subseteq C$ , where  $U = \{x_1, x_2, \dots, x_n\}$  is a finite set of samples,  $C$  is the condition attribute set and  $D = \{d\}$  is the decision attribute set.  $B$  is a real-valued attribute subset for describing the samples. The  $\delta$ -neighborhood of an arbitrary sample

$x_i \in U$  is defined as follows.

$$\delta_B(x_i) = \{x_j | x_j \in U, \Delta_B(x_i, x_j) \leq \delta\} \tag{1}$$

where  $\delta_B(x_i)$  is also called the  $\delta$ -neighborhood information granule of  $x_i$ .  $\delta$  is the neighborhood threshold, which is a real number between 0 and 1.  $\Delta_B(x_i, x_j) = \left(\sum_{v \in B} |f(x_i, v) - f(x_j, v)|^2\right)^{1/2}$  is the distance function. Formula (1) represents sample set with similar attribute values to those of sample  $x_i$  for attribute set  $B$ . The smaller  $\delta$  is, the higher the similarity between sample  $x_i$  and sample  $x_j$  in the attribute space  $B$ , and the smaller the number of individuals in the neighborhood. When  $\delta = 0$ , the neighborhood rough set becomes the classical rough set. Furthermore, the sample set is divided into sample set  $d_0$  of the increasing drum height, sample set  $d_1$  of the decreasing drum height, and sample set  $d_2$  of the holding drum height according to the decision attribute  $d$ . The positive region of the sample set with respect to the feature set  $B$  is defined as follow.

$$POS_B(d) = \underline{N}_B^\delta(d_0) \cup \underline{N}_B^\delta(d_1) \cup \underline{N}_B^\delta(d_2) \tag{2}$$

where  $\underline{N}_B d_0 = \{x_j | \delta_B(x_j) \subseteq d_0, x_j \in U\}$ ,  $\underline{N}_B d_1 = \{x_j | \delta_B(x_j) \subseteq d_1, x_j \in U\}$ ,  $\underline{N}_B d_2 = \{x_j | \delta_B(x_j) \subseteq d_2, x_j \in U\}$ . According to formula (2), all samples with similar feature values to a sample which is in the positive region, have the same classification. This shows that the sample can be accurately classified under the feature set  $B$ . The more samples there are in the positive domain, the better the separability of the attribute set and the stronger the classification ability, which is beneficial to the learning of the classification algorithm. According to the decision information system of the shearer drum height regulating, neighborhood rough set-based attribute reduction is introduced to generate a set of reducts, and then each reduct is used to train a base classifier. To achieve this goal, a forward feature selection strategy is used based on attribute importance.

$$\gamma_B^\delta(d) = \frac{|POS_B(d)|}{|U|} \tag{3}$$

where  $|\cdot|$  is the cardinality of the set.  $\gamma_B^\delta$  is the  $\delta$ -dependency of  $d$  on  $B$ .

The significance of the attribute  $c$  in  $B$  relative to the decision  $d$  is as (4).

$$SIG(c, B, d) = \gamma_{B \setminus c}^\delta(d) - \gamma_B^\delta(d) \tag{4}$$

To obtain different reducts, we generate  $k$  different neighborhood radii in  $[\delta_{min}, \delta_{max}]$ . By setting different neighborhood thresholds  $\delta_i$ , we can control the granularity size of the sample, so as to obtain the reduction results at different levels. Furthermore, we can get  $p$  different sample sets  $S = \{S_1, S_2, \dots, S_p\}$  according to the reducts. The sample set  $S$  is divided into training set  $M = \{M_1, M_2, \dots, M_p\}$  and test set  $T = \{T_1, T_2, \dots, T_p\}$ .

**C. OSL-KELM-SDHR BASE CLASSIFIER**

Based on the advantages of the KELM with high learning speed, an online self-learning extreme learning machine of the shearer drum height regulating (OSL-KELM-SDHR) classifier is proposed, which is used to combine an ensemble classifier for adjusting cutting height of the shearer drum. In this section, the OSL-KELM-SDHR model is discussed in detail.

1) ELM-SDHR BASE CLASSIFIER

Set the training samples of field data  $M = \{(\mathbf{x}_i, T_i) \mid i = 1, 2, \dots, N\}$ , where  $\mathbf{x}_i \in \mathbf{R}^d$  is a  $d$ -dimensional input vector and  $T_i \in \mathbf{R}$  is the output value corresponding to  $\mathbf{x}_i$  (different values represent different adjustment patterns). The ELM model is defined as follow [31].

$$\begin{cases} \min_{\beta, \xi} \frac{1}{2} \|\beta\|^2 + c \frac{1}{2} \sum_{i=1}^t \xi_i^2 \\ \text{s.t. } h(\mathbf{x}_i)\beta = y_i - \xi_i \end{cases} \quad (5)$$

where  $\beta = [\beta_1, \beta_2, \dots, \beta_L]^T \in \mathbf{R}^{L \times N}$  is the output weight vector of the model,  $\mathbf{h}(\mathbf{x}_i) = [h_1(\mathbf{x}_i), h_2(\mathbf{x}_i), \dots, h_L(\mathbf{x}_i)]$  is the attribute mapping from the  $d$ -dimensional input to  $L$ -dimensional hidden layer attributes,  $\xi_i$  is the training error corresponding to the  $i$  th sample and  $c$  is the regularization parameter,  $c \in \mathbf{R}^+$ .

Based on KKT optimization conditions, the optimization problem of formula (5) is solved, and the weight vector parameters  $\hat{\beta}_{BELM}$  of the shearer shearing path adjustment model are calculated.

$$\hat{\beta}_{BELM} = \mathbf{H}^T \left( \frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T} \quad (6)$$

where  $\mathbf{T} = [T_1, T_2, \dots, T_N]^T$  is the target value vector of the input samples, and  $\mathbf{H}(\mathbf{x}_i) = [h^T(\mathbf{x}_1), h^T(\mathbf{x}_2), \dots, h^T(\mathbf{x}_N)]^T$  is the mapping matrix of the input samples.

Assuming that the input vector of a new sample is  $\mathbf{x}_p$ , the height regulating mode  $T_{scp}$  of the shearer drum can be calculated by formula (7).

$$T_{scp} = \mathbf{h}(\mathbf{x}_p) \hat{\beta}_{BELM} \quad (7)$$

2) KELM-SDHR BASE CLASSIFIER

To further improve the accuracy of the base classifier and enhance the generalization performance with complex coal seam, we propose a KELM-SDHR classifier model based on KELM. Given a new input vector  $\mathbf{x}_p$ , the output  $T_{scp}$  is calculated according to the weight vector parameter  $\hat{\beta}_{RELM}$ .

$$T_{scp} = \mathbf{h}(\mathbf{x}_p) \hat{\beta}_{RELM} = \mathbf{h}(\mathbf{x}_p) \mathbf{H}^T \left( \frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T} \quad (8)$$

According to the theory of the inner product of the kernel function [32], we can directly use the kernel function instead of the nonlinear explicit mapping HHT of the ELM hidden

layer nodes to improve the robustness and generalization ability of the ELM.

$$T_{scp} = \mathbf{K}\mathbf{A}^{-1}\mathbf{T} \quad (9)$$

where  $K(\mathbf{x}_i, \mathbf{x}_j)$  is the kernel function,  $\mathbf{A}^{-1} = \left( \frac{\mathbf{I}}{C} + \mathbf{\Omega}_{KELM} \right)^{-1}$  and  $\mathbf{K} = [K(\mathbf{x}_p, \mathbf{x}_1), \dots, K(\mathbf{x}_p, \mathbf{x}_N)]$ . Formula (9) is the classification model of the KELM-SDHR. The model does not need to provide an explicit hidden layer attribute mapping function  $h(x)$  and the number of hidden layer neurons  $L$ .

3) ONLINE SELF-LEARNING KELM-SDHR BASE CLASSIFIER

The boundary of a coal seam is complex and changeable in a fully mechanized coalface, but there is law. With the advancement of the working face, the newest structure information of the coal seam is exposed. However, the KELM-SDHR is unable to learn the newly acquired structure information of the coal seam online. With time passing by, the accuracy of the KELM-SDHR may decline. So, it is necessary to provide a dynamic update mechanism to improve the prediction accuracy of the KELM-SDHR. In view of this, this paper proposes a dynamic OSL-KELM-SDHR model based on the KB-IELM [33]. The data with labels are recorded by a remote control system in real time while operating the shearer. The label of new samples is based on the manual control and ensemble controller. The main steps using the OSL-KELM-SDHR are as follows.

Step 1: Initialization:

A classifier  $T_t(\mathbf{x}) = \mathbf{K}_t \mathbf{A}_t^{-1} \mathbf{T}_t$  is trained according to the initial samples  $M_t = \{\mathbf{x}_i, T_i\}_{i=1}^{N_0}$ , where  $N_0$  is the number of initial samples,  $\mathbf{K}_t = [K(\mathbf{x}, \mathbf{x}_1), \dots, K(\mathbf{x}, \mathbf{x}_{N_0})]$ , and

$$\begin{aligned} \mathbf{A}_t^{-1} &= \left( \frac{1}{C} \mathbf{I}_t + \mathbf{\Omega}_{KELM}^t \right)^{-1} \\ &= \left( \begin{array}{ccc} \frac{1}{C} + K(\mathbf{x}_1, \mathbf{x}_1) & \dots & K(\mathbf{x}_1, \mathbf{x}_{N_0}) \\ \vdots & \ddots & \vdots \\ K(\mathbf{x}_{N_0}, \mathbf{x}_1) & \dots & \frac{1}{C} + K(\mathbf{x}_{N_0}, \mathbf{x}_{N_0}) \end{array} \right)^{-1} \end{aligned}$$

Step 2: Online self-learning. Repeat the following process:

It is assumed that  $k$  new samples with labels are obtained in the working process of the shearer. A new sample set with the initial samples and the new samples is formed. Then, we use the new sample set to train a new classifier.

$$\begin{aligned} T_{t+1}(\mathbf{x}) &= \mathbf{K}_{t+1} \mathbf{A}_{t+1}^{-1} \mathbf{T}_{t+1} \\ &= (\mathbf{K}_t \quad \Delta \mathbf{K}_{t+1}) \begin{pmatrix} \mathbf{A}_t & \mathbf{U}_t \\ \mathbf{U}_t^T & \mathbf{D}_t \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{T}_t \\ \Delta \mathbf{T}_{t+1} \end{pmatrix} \end{aligned} \quad (10)$$

where  $\mathbf{K}_{t+1} = [K(\mathbf{x}, \mathbf{x}_{N_0+1}), \dots, K(\mathbf{x}, \mathbf{x}_{N_0+k})]$  is kernel mapping of  $\mathbf{x}$  for newly added training samples,

$$\mathbf{U}_t = \begin{pmatrix} K(\mathbf{x}_1, \mathbf{x}_{N_0+1}) & \dots & K(\mathbf{x}_1, \mathbf{x}_{N_0+k}) \\ \dots & \ddots & \dots \\ K(\mathbf{x}_{N_0}, \mathbf{x}_{N_0+1}) & \dots & K(\mathbf{x}_{N_0}, \mathbf{x}_{N_0+k}) \end{pmatrix},$$

$$D_t = \begin{pmatrix} \frac{1}{C} + K(\mathbf{x}_{N_0+1}, \mathbf{x}_{N_0+1}) & \dots & K(\mathbf{x}_{N_0+1}, \mathbf{x}_{N_0+k}) \\ \dots & \ddots & \dots \\ K(\mathbf{x}_{N_0+k}, \mathbf{x}_{N_0+1}) & \dots & \frac{1}{C} + K(\mathbf{x}_{N_0+k}, \mathbf{x}_{N_0+k}) \end{pmatrix}.$$

Using the block matrix inverse formula [34] with  $A_{t+1}^{-1}$ , we can obtain

$$A_{t+1}^{-1} = \begin{pmatrix} A_t^{-1} + A_t^{-1} U_t \rho_t^{-1} U_t^T A_t^{-1} & -A_t^{-1} U_t \rho_t^{-1} \\ -\rho_t^{-1} U_t^T A_t^{-1} & \rho_t^{-1} \end{pmatrix} \quad (11)$$

where  $\rho_t = D_t - U_t^T A_t^{-1} U_t$ .

The incremental updating basis for the classification model OSL-KELM-SCT is obtained by incorporating formula (11) into formula (10).

$$\begin{aligned} T_{t+1}(\mathbf{x}) &= K_{t+1} A_{t+1}^{-1} T_{t+1} \\ &= (K_t \quad \Delta K_{t+1}) \\ &\quad \times \begin{pmatrix} A_t^{-1} + A_t^{-1} U_t \rho_t^{-1} U_t^T A_t^{-1} & -A_t^{-1} U_t \rho_t^{-1} \\ -\rho_t^{-1} U_t^T A_t^{-1} & \rho_t^{-1} \end{pmatrix} \\ &\quad \times \begin{pmatrix} T_t \\ \Delta T_{t+1} \end{pmatrix} \\ &= K_t A_t^{-1} T_t + K_t A_t^{-1} U_t \rho_t^{-1} U_t^T A_t^{-1} T_t \\ &\quad - \Delta K_{t+1} \rho_t^{-1} U_t^T A_t^{-1} T_t \\ &\quad - K_t A_t^{-1} U_t \rho_t^{-1} \Delta T_{t+1} + \Delta K_{t+1} \rho_t^{-1} \Delta T_{t+1} \\ &= T_t(\mathbf{x}) + (\mathbf{B} - \Delta K_{t+1}) \rho_t^{-1} (\mathbf{E} - \Delta T_{t+1}) \end{aligned} \quad (12)$$

where  $\mathbf{B} = K_t A_t^{-1} U_t$  and  $\mathbf{E} = U_t^T A_t^{-1} T_t$ .

Step 3: Online use

The application of the OSL-KELM-SDHR includes two stages.

(1) Self-learning optimization of the basic classifier. Considering the working cycle of the shearer and the structure of the coal seam, the learning cycle and the number of newly added samples are determined. Assuming that the learning time period is  $t$ , the  $N_t$  samples are obtained. The newer the samples, the better generalization the trained classifier will get due to the local sudden change of the coal seam. Therefore, the forgetting mechanism is needed to enhance the impact of new samples on the classifier and remove the old samples. Assume that  $N_{max}$  is the maximum number of training samples in the model. The OSL-KELM-SDHR discards  $d$  samples according to the historical timing of the samples when the model is updated in the  $(m+1)$ th in order to guarantee the number of samples  $N_0 + \sum_{i=1}^{m+1} N_i - d \leq N_{max}$ .

(2) Decision-making using the basic classifier. Suppose that the total number of samples is  $M \leq N_{max}$  after several updates, and then these samples are used to calculate the parameter  $A_M^{-1}$ . When the input sample is  $\mathbf{x}_{test}$ , the shearer drum performs the action according to  $T_{test} = [K(\mathbf{x}_{test}, \mathbf{x}_1), \dots, K(\mathbf{x}_{test}, \mathbf{x}_M)] \cdot A_M^{-1} [T_1, T_2, \dots, T_M]^T$

#### D. SELECTIVE ENSEMBLE

The base classifiers trained with reducts and OSL-KELM-SDHR algorithm will have a good generalization power. What is more, the classifiers trained with different reducts should be diverse. It seems a good solution to construct ensemble model with the neighborhood rough set-based reducts. However, with experimental analysis, we find that there are usually hundreds of reducts for the field data. This triggers a problem: whether we require using all the reducts or not. If a small number of base classifiers are combined, the difference of the base classifiers is difficult to fully reflect, and the ensemble classification accuracy is insufficient. If all base classifiers are combined, the classification accuracy also will decrease due to the characteristics of the ensemble learning. However, it still has been difficult in how to choose base classifiers [35]. Although a large number of metrics and selection algorithms have been proposed, there is no consistently selective ensemble method that is better than other technologies in engineering applications [36]. Therefore, we introduce an accuracy-guided forward search and post-pruning strategy FS-PP [26] which selects part of the base classifiers to construct an efficient and effective ensemble system for the shearer drum lifting prediction.

### III. EXPERIMENTS AND RESULTS

#### A. FIELD DATASET

##### 1) OBTAINING THE ORIGINAL DATA SET

The dataset was collected from the 43101 fully mechanized coalface of the Yujialiang Coal Mine in northern Shaanxi, China. The length of the 43101 working face is 351.4 m, the thickness of the coal seam is 1.0-1.7 m, the average thickness is 1.47 m, the average height of the working faces is 1.4 m, and the inclination is 3°-5°. The main configuration of the fully mechanized mining equipment is as follows: The hydraulic support adopts the ZY9200/09/18D double-pillar shield type with a supporting strength of 0.99-1.06 MPa; the shearer adopts the MG 2 × 200.890-WD1 type, the shearer shearing height range is 1.3-2.5 m, and the total installed power is 890 kW; the scraper selects the SGZ800/1400 type, and the installed power is 2 × 700 kW. The layout structure of the 43101 fully mechanized mining faces in the Yujialiang Coal Mine is shown in Fig 3. The time interval for collection is from March 29 to 30, 2020, from 01:00 am to 06:00 am each day, and the sampling period is 1.0 s. The raw data set is composed of 26 conditional attributes and 1 decision attribute. The total number of samples is 26,343 after removing duplicate samples.

##### 2) DATA PREPROCESSING

The prediction of the adjustment mode of a shearer shearing drum is a complex classification problem based on data mining. The important premise behind machine learning theory is that we must have enough sample data with high quality. The quality of the data determines the quality of the model extracted by the classification algorithm or trained. When the sample quality is poor, it is difficult for the proposed

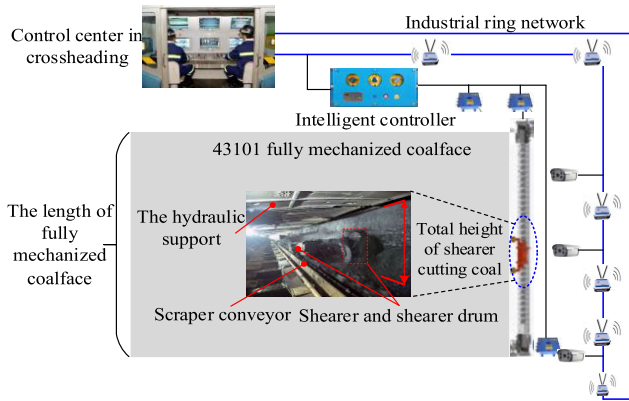


FIGURE 3. 43101 fully mechanized coalface.

model to achieve the expected results. Therefore, in the data preprocessing stage, the raw data set is processed to remove missing values and outliers, and then the data are normalized to construct a high-quality data set.

There are some data missing in the original sample data, including one-dimensional and multidimensional data. For samples with missing one-dimensional data, this article uses Lagrangian interpolation to impute the missing values; for samples with missing multidimensional data, to ensure accuracy, the sample is deleted. Some shearer shearing drum adjustment mode records are outliers; due to electrical failure or sensor failure, encountering some abnormal values is inevitable. According to the actual characteristics of these abnormal data, they are treated as missing values or corrected by using average values, and some are deleted.

To eliminate the dimensional interactions between the attributes, the original sample data are subjected to the minimum-maximum normalization process [37], where the relationship existing in the original data is retained, and each attribute value is mapped to [0,1] to achieve normalization. After preprocessing and normalization, the samples have 23,827 samples. Among the output variables, there are 6995 data in which the shearer drum rises, 7728 data in which the shearer drum falls, and 9104 data where the height remains the same. The data set composed of the attribute set  $\{c_1-c_{14}, d\}$  is called data set I, and the data set composed of the attribute set  $\{c_1-c_{26}, d\}$  is called data set II. The statistical results of the output variables in the two data sets are shown in Fig 4. The sample distribution has a small proportion of imbalance.

### 3) DATASET ESTABLISHMENT

This study uses a 10 folds cross-validation method to evaluate the performance of the proposed model. The attribute subsets are obtained based on the field data. According to each attribute subset, a corresponding sample subset is generated, and the number of samples is the same as that of the original data sample (duplicate samples are not removed). The samples split into a training subset of 70% and a subtest set with the remaining 30% samples. The whole training (testing)

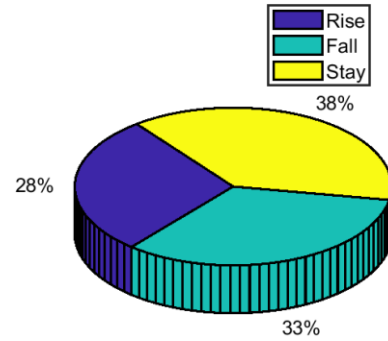


FIGURE 4. Statistics of the data.

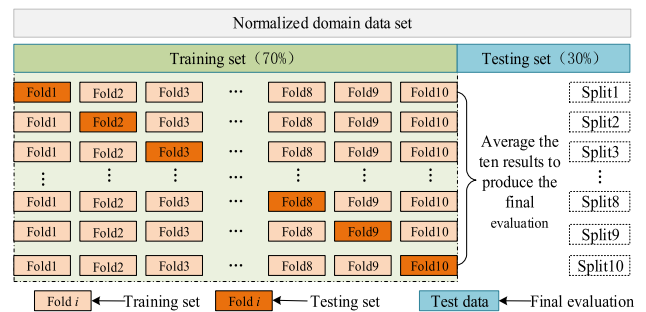


FIGURE 5. Tenfold cross validation.

set is composed of multiple training (subtest) subsets, and the training (subtest) sets depends on the attribute reduction subsets. All training subsets were partitioned into 10 folds using cross-validation (one fold for test 9 folds for validating), and this is done on each dataset for 10 times, that is, each classifier is trained for 10 times (Fig 5.). The validating sets in all training sets are used to obtain the optimal hyperparameters for the base classifiers.

### B. PERFORMANCE MEASURES

To accurately and comprehensively evaluate the ensemble classification model for shearer shearing drum height regulating mode, four evaluation metrics have been used: accuracy, precision, recall, and F1-scores [38] which are the most popular metrics to measure the performance of a classifier. A confusion matrix helps to intuitively compare and understand the classification performances of the proposed model. In the confusion matrix, predicted categories are represented by rows, and actual categories are represented by columns. Table 2 is the confusion matrix for classifications of the shearer shearing drum height regulating mode, where U, D and S represent increasing its height, decreasing its height and holding. UU, DD, and SS represent the number of samples with correct predictions, and other combinations represent the number of samples with incorrect predictions.

Based on the confusion matrix, the classification performance evaluation index of this study is defined as follows.

$$Accuracy = \frac{UU + DD + SS}{p} \tag{13}$$

where  $p$  is the total number of test samples.



**TABLE 2. Confusion matrix for height prediction of shearer drum.**

Actual	Predicted		
	U	D	S
U	UU	UD	US
D	DU	DD	DS
S	SU	SD	SS

Generally, the precision and recall are applied together in the F-measure for assessing the prediction accuracy of a classification model. These measures for height prediction problem of shearer drum are calculated as follows.

$$Precision = \frac{1}{3} \left( \frac{UU}{UU + DU + SU} + \frac{DD}{DD + UD + SD} + \frac{SS}{SS + US + DS} \right) \quad (14)$$

$$Recall = \frac{1}{3} \left( \frac{UU}{UU + UD + US} + \frac{DD}{DD + DU + DS} + \frac{SS}{SS + SU + SD} \right) \quad (15)$$

$$F1 - Score = \frac{2 Precision \times Recall}{Precision + Recall} \quad (16)$$

### C. EXPERIMENTAL SETUP

#### 1) LAB ENVIRONMENT

In this study, we use MATLAB R2018b for program development and perform our experiments on a hardware platform with an Intel(R) Core(TM) i5-6500 CPU @ 3.20 GHz, 16.0 GB RAM, and the Windows 7 64-bit operating system.

#### 2) MODEL PARAMETERS

The parameters of the proposed model in this paper mainly include two categories, one is the neighborhood radii of the neighborhood rough set attribute reduction (NRS-AR) and the number of seeds  $K$ , and the other is the parameters of the base classifier ELM-SDHR and KELM-SDHR. We obtain the attribute subsets by changing the value  $K$  in the numerical range  $[0.1, 0.6]$  [39]. The initial value  $K$  is 50. All the hyperparameters of the base classifiers are selected with 10-fold cross validation while the base classifiers with the highest accuracies. The hyper-parameters for different classifiers in the proposed model are described in Table 3.

#### 3) COMPARISON APPROACH

To evaluate the proposed method, three different perspectives are selected for comparison with the proposed SDHRM-SEoKELMRS:

(1) Three base classifiers are listed in Table 4. Currently, there are many learning algorithms that can be used as basic classifiers for ensemble classification. However, we select three different categories of the most popular classification algorithm to compare with the base classification algorithm

**TABLE 3. Hyperparameters adopted for the models.**

Model	Hyperparameters	Empirical scope	Final value
NRS-AR	Neighborhood radius $\delta$	[0.1,0.6]	[0.1,0.6]
	Seed number $K$	[50,200]	150
ELM	Number of Hidden Units	[3,100]	20
	Activation function	Sigmoid(),tanh(), Relu(),et al.	Sigmoid()
	The weight between input and hide layer $w$	-	random-number
	Biases of hidden neural units $b$	-	random-number
KELM	Number of hidden units	[3,100]	20
	Kernel function	Linear kernel, Polynomial kernel, Gauss kernel, et al.	Gauss kernel
	Regularization parameter $C$	-	$1.5 \times 10^4$
	The width of Gaussian kernel $\delta$	-	10

**TABLE 4. Base classifiers and hyperparameters.**

Type	Learning Algorithm	Hyperparameters	Final value
Lazy	KNN	Number of Neighbors	7
		Distance	Euclidean
Tree	CART	SplitCriterion	gdi
		MinLeafSize	53
		MaxNumSplits	13,826
Functions	SVM	Coefficient of the penalty term ( $c$ )	6.4772
		Gamma value of Gaussian kernel ( $\gamma$ )	5.6063

selected in this article. All these algorithms are imported from the classification learning algorithm package of MATLAB R2018b. These algorithms are widely used for ensemble classification problems, which are simple and effective. There is no unified theory or guiding method for the selection of the related hyperparameters in the algorithm listed in Table 4. This paper uses a 10-fold cross-validation experiment to determine the optimal hyperparameters of each base classifier.

(2) Two traditional integrated classifiers are used: the bagging and random subspace methods. To show the superiority of the proposed FS-PP ensemble classification method, we use the bagging and random subspace methods to establish the same number of base classifiers as in the FS-PP and integrate them to carry out related comparative experiments.

(3) Two original attribute sets are used: the original attribute set containing transfer attributes and the original attribute set without transfer attributes. To show the influence of migration attributes on the classification effect for shearer shearing drum adjustment modes, the FS-PP ensemble

**TABLE 5. Results of reduction.**

	Data	Attributes	Reducts
1	Without migration attributes(Data set I)	14	85
2	Containing migration attributes(Data set II)	26	128

classifier proposed in this paper is selected to compare the classification effects between the two original attribute sets.

Our comparison method is as follows: In Section D.1, we show the neighborhood data reduction subset based on the domain-granulated rough set; in Section D.2, we report the classification results of different base classifiers on the domain data set; in Section D.3, the relationship between the number of base classifiers and different ensemble classification performance indicators is shown; finally, in Section D.4, the ensemble classification method proposed in this paper is compared with the traditional ensemble classification method, and the classification results of the integrated system under the two sets of attribute spaces are presented.

## D. RESULTS AND DISCUSSION

### 1) ATTRIBUTE SUBSET

In the neighborhood threshold interval  $[0.1, 0.6]$ , by adjusting the number of granulated seeds  $K$  in the neighborhood, Algorithm [26] is used to perform attribute reduction to the field data set. To ensure the complementarity of the classification information from attribute reduction subsets, we select as many attribute subsets as possible. In addition, to ensure the diversity of the base classifiers, each group of attribute subsets should have attributes that are at least 30% different. Combining the above two points, Table 5 shows the final numbers of reduced subsets that do not include transfer attributes and those that include transfer attributes.

### 2) THE PERFORMANCE OF SINGLE CLASSIFIERS

According to Table 4, the optimal parameters are described for each classification algorithm in Section Table 3 and Table 4, and the classification effect of each algorithm on the shearer shearing drum height adjustment data set is analyzed. The base classifiers listed in Table 3 are independently evaluated using the field data set (data set I does not contain migration attributes and data set II contains migration attributes), and each base classifier is shown in Table 7 and Table 8, which show the accuracy, precision, recall and F1-score results. Among them, Table 6 contains the classification prediction results for the data set that does not contain migration attributes, and Table 7 contains the classification prediction results for the data set that contains migration attributes.

Table 6 and Table 7 summarize the prediction performances of different classifiers on the two sets of data. It can be seen that the KELM has the best prediction performance on the two data sets, and the highest prediction accuracy rates for the two data sets are 71.99 and 73.21, respectively. However, the prediction performance of the ELM on the two

**TABLE 6. The performance of base classifiers for data set I.**

Metric	Classifier				
	KNN	CART	SVM	ELM	KELM
Accuracy	69.14	66.59	70.13	57.10	71.29
Precision	67.41	66.57	68.92	48.78	70.69
Recall	66.57	67.63	67.32	53.48	69.63
F1-Score	66.52	67.65	68.14	51.02	70.16

**TABLE 7. The performance of base classifiers for data set II.**

Metric	Classifier				
	KNN	CART	SVM	ELM	KELM
Accuracy	71.38	69.47	71.11	58.81	72.21
Precision	68.12	66.80	71.38	62.78	70.73
Recall	69.35	65.76	70.32	56.76	68.48
F1-Score	68.73	65.98	70.85	59.52	69.59

sets of data is poor, and the prediction accuracy rates are 57.10 and 58.81, respectively. Among the five classification algorithms, the classification performances of the SVM and KELM are relatively close. The classification performances of different classification algorithms on data sets that contain migration attributes are better than those of data sets that do not contain the status transfer attributes. The prediction accuracy is improved by 1.80% on average, and the prediction accuracy of the CART algorithm is improved by up to 2.98%.

### 3) THE NUMBER OF BASE CLASSIFIERS

The number of base classifiers determines the final generalization performance of the ensemble algorithm. As the number of base classifiers increases, although the ensemble classification performance is improved to a certain extent, the data scale and the number of integrated classifiers increase, and the need to consume many computing resources causes the classification performance to decline. In this experiment, first, on the basis of the original data set I and the original data set II, 85 attribute subsets that do not contain transfer attributes and 128 attribute subsets that contain transfer attributes generated by reduction are used to generate the training data set of the base classifier. Then, the generated data set is applied to five classification algorithms, the KNN, CART, SVM, ELM, and KELM methods, to obtain two groups of base classifiers. Each group contains 5 types of base classifiers, each with 85 and 128 attribute subsets in dataset I and dataset II, respectively. Finally, the selective ensemble strategy in algorithm 2 is used to determine the number of base classifiers from each group to be finally integrated. The effect of the number of selective ensemble base classifiers on the performance of the ensemble system in terms of classification accuracy is shown in Fig 6.

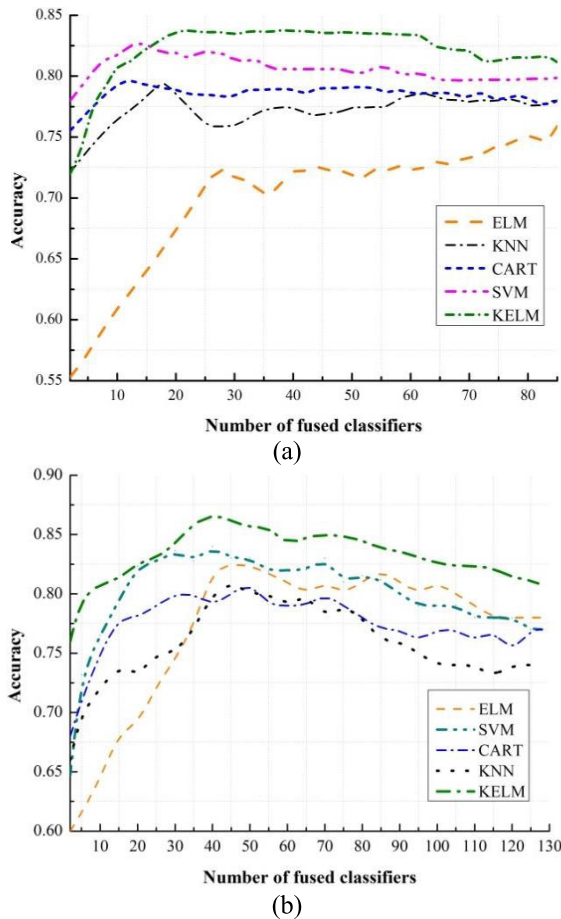


FIGURE 6. Variation of classification accuracies with number of base classifiers. (a) Dataset I. (b) Dataset II.

Fig 6 shows that as the number of base classifiers increases, the recognition accuracy of the shearer drum height adjustment mode first gradually increases and then gradually decreases. The reason why the ensemble classification accuracy rate gradually rises during the initial stage of the ensemble is that at the beginning of the ensemble sorting process, the base classifier included in the ensemble is small, and the difference between the base classifiers is small, which directly leads to low ensemble accuracy. While continuously adding new base classifiers with better comprehensive performance, the diversity of the base classifiers is increased, so the ensemble accuracy gradually increases. After the ensemble accuracy reaches a certain value, the ensemble accuracy gradually decreases because at this time, after adding enough base classifiers, a base classifier with poor comprehensive performance appears in the integrated system, and the ensemble accuracy gradually decreases. In dataset I, which does not contain migration attributes, when the KELM base classifier is between 20 and 25, the recognition of the shearer drum height adjustment pattern can be highly accurate. However, as the number of base classifiers increases further (the number of base classifiers is greater than 60), the performance in terms of recognizing the shearer drum height adjustment pattern is not very helpful, and even the addition of too

TABLE 8. Number of fused different classifiers.

Data set	Classifiers	Number of fused classifiers
I	ELM	28
	KNN	14
	CART	12
	SVM	15
	KELM	20
II	ELM	42
	KNN	42
	CART	35
	SVM	25
	KELM	40

TABLE 9. Comparison of the performance of ensemble classifiers (ELM).

Data set	Metric	Single Classifiers (best)	Traditional Ensembles (best)		Our approach FS-PP
			Bagging	Random Subspace	
I	Accuracy	68.61	79.58	76.89	81.12
	Precision	67.47	80.81	78.33	79.58
	Recall	74.70	79.56	79.65	79.40
	F1-Score	70.90	75.47	78.98	79.49
	Test time	0.259	0.925	1.159	0.931
II	Accuracy	70.02	83.94	79.18	82.42
	Precision	90.63	80.06	78.67	78.67
	Recall	79.94	75.28	75.29	74.70
	F1-Score	84.95	77.60	76.94	76.63
	Test time	0.306	1.065	1.278	0.957

many weak learners leads to a decrease in the final accuracy rate. Therefore, in subsequent comparative experiments, the final ensemble base classifiers number is 20 for data set I. In dataset II, which contains migration characteristics, when the KELM base classifier is between 35 and 45, the recognition of the shearer drum height mode can be highly accurate, and the number of base classifiers is determined to be 40. The ensemble numbers of other base classifiers is shown in Table 8.

4) COMPARISON BETWEEN SDHRM-SEoKELMRS AND OTHERS CLASSIFIER ENSEMBLE

Finally, we compare and analyze the classification performance of the SDHRM-SEoKELMRS along with those traditional ensemble classifier, the bagging and random subspace methods, on data set I and on data set II. The traditional ensemble classification methods (the bagging and random subspace methods) and the SDHRM-SEoKELMRS algorithm contain the same number of base classifiers. The test set is used to predict the shearer drum height regulating pattern shown in Fig 5. The base classifier algorithms adopt the ELM, KNN, CART, SVM, KELM, and OSL-KELM. Tables 9, 10, 11, 12, 13, and 14 show the results of the experiments.

Tables 9-14 summarize the prediction performance of the base classifier on the test data set containing shearer drum height adjustment modes under different ensemble methods.

**TABLE 10. Comparison of the performance of ensemble classifiers (KNN).**

Data set	Metric	Single Classifiers (best)	Traditional Ensembles (best)		Our approach
			Bagging	Random Subspace	FS-PP
I	Accuracy	69.39	80.23	74.54	78.13
	Precision	72.54	78.01	77.69	78.52
	Recall	70.28	77.76	76.58	76.48
	F1-Score	71.39	77.88	77.13	77.49
	Test time	0.546	1.850	2.545	2.122
II	Accuracy	71.43	79.41	77.59	80.36
	Precision	76.18	80.48	79.87	83.36
	Recall	72.69	77.89	77.53	79.67
	F1-Score	74.88	79.16	78.68	81.47
	Test time	0.618	2.153	2.974	2.484

**TABLE 11. Comparison of the performance of ensemble classifiers (CART).**

Data set	Metric	Single Classifiers (best)	Traditional Ensembles (best)		Our approach
			Bagging	Random Subspace	FS-PP
I	Accuracy	71.58	79.36	78.13	81.91
	Precision	78.04	78.18	81.53	80.97
	Recall	73.00	78.04	78.30	73.29
	F1-Score	75.44	78.11	80.86	76.94
	Test time	0.225	0.919	1.525	1.208
II	Accuracy	76.47	82.02	81.25	83.74
	Precision	85.28	80.78	82.39	85.61
	Recall	72.53	90.00	76.06	87.64
	F1-Score	78.39	85.14	79.10	86.61
	Test time	0.367	1.681	2.312	2.273

**TABLE 12. Comparison of the performance of ensemble classifiers (SVM).**

Data set	Metric	Single Classifiers (best)	Traditional Ensembles (best)		Our approach
			Bagging	Random Subspace	FS-PP
I	Accuracy	65.38	80.45	78.38	82.47
	Precision	72.55	80.95	83.97	87.78
	Recall	66.67	73.02	84.13	88.89
	F1-Score	69.49	76.78	84.05	88.33
	Test time	0.652	1.931	3.140	2.875
II	Accuracy	69.70	82.66	81.70	85.37
	Precision	75.80	80.42	82.11	81.55
	Recall	61.75	71.43	76.19	80.95
	F1-Score	68.06	75.66	79.04	81.25
	Test time	0.855	2.312	4.102	3.531

The prediction accuracy of the SDHRM-SEoKELMRS method on the two test sets (83.78 and 85.68, respectively) is higher than the prediction accuracies of other base classifiers, which are comparable to that of a single OSL. Compared with the KELM-based classifier, the accuracy of the ensemble method is increased by 15.60% and 13.86%, respectively. The prediction accuracy of each prediction model on test set II is generally higher than that on test set I. We believe that this shows the effectiveness of using the migration attribute

**TABLE 13. Comparison of the performance of ensemble classifiers (KELM).**

Data set	Metric	Single Classifiers (best)	Traditional Ensembles (best)		Our approach
			Bagging	Random Subspace	FS-PP
I	Accuracy	72.25	81.51	78.90	81.03
	Precision	87.25	87.39	86.67	86.04
	Recall	66.11	87.55	74.54	78.48
	F1-Score	75.23	87.75	80.15	82.09
	Test time	0.447	1.395	2.244	2.095
II	Accuracy	74.63	81.56	79.84	82.87
	Precision	87.88	85.12	85.98	88.99
	Recall	68.89	76.25	71.98	82.23
	F1-Score	77.23	80.44	78.36	85.48
	Test time	0.578	1.535	2.516	2.317

**TABLE 14. Comparison of the performance of ensemble classifiers (OSL-KELM).**

Data set	Metric	Single Classifiers (best)	Traditional Ensembles (best)		Our approach
			Bagging	Random Subspace	FS-PP
I	Accuracy	72.47	84.02	80.83	83.78
	Precision	81.44	83.69	81.64	82.87
	Recall	73.52	83.76	83.01	81.47
	F1-Score	77.28	83.73	82.32	83.17
	Test time	0.261	0.683	1.142	1.083
II	Accuracy	75.25	83.91	82.93	84.68
	Precision	82.73	85.18	81.06	84.08
	Recall	81.81	84.33	80.74	85.89
	F1-Score	82.26	84.73	80.90	84.97
	Test time	0.278	0.786	1.335	1.230

set to improve the accuracy of the classification prediction for the shearer drum. The proposed SDHRM-SEoKELMRS ensemble classifier method achieves the best predictive performance for these four metrics (accuracy = 85.68, precision = 84.08, recall = 85.89 and F1-Score = 84.97) on test set II. The FS-PP and random subspace methods both obtain multiple different groups of base classifiers by perturbing the sample attribute space, but for the industrial test data set of shearer drum height adjustment modes, the FS-PP method uses the redundancy of the neighborhood rough set attribute reduction technology to effectively improve the generalization performance of the integrated system.

To better compare the prediction accuracy and calculating time of the classification algorithm and its integrated system on the industrial test set of shearer drum height adjustment modes, we provide the results in the form of graphs, as shown in Fig 7 and Fig 8. Although our method does not achieve a completely accurate prediction effect, the obtained results are acceptable in comparison with those of other approaches. Although the response time of the

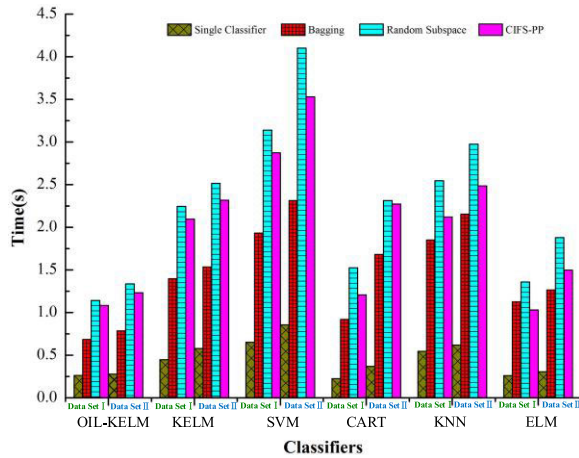


FIGURE 7. Times comparison between ensemble classifier.

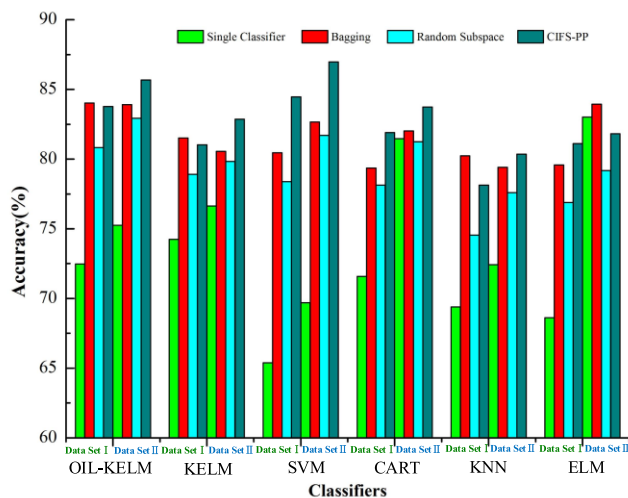


FIGURE 8. Accuracy comparison between ensemble classifiers.

FS-PP +ELM combination is as short as 0.957 s, the prediction accuracy rate drops by 4.72%. The FS-PP+SVM combination has the highest prediction accuracy rate at 86.97, but the long amount of time required to make a decision makes it difficult to meet real-time performance requirements. In the process of industrial production, the actual operating speed of the shearer is between 0-0.08 m/s. In view of the decision-making time and prediction accuracy, the proposed SDHRM-SEoKELMRS method has the best performance. In addition, we can also adjust the maximum number of training samples of the SDHRM-SEoKELMRS algorithm or optimize the adaptive calculation method for the decision-making parameter to further enhance the real-time decision-making of the algorithm to adapt it to the needs of working conditions. This is also the focus of our future work. The SDHRM-SEoKELMRS method adopts an incremental learning method, which enhances the adaptability of the shearer to the randomly changing coal roof in the coal mine and can effectively predict the adjustment mode of the drum

to avoid the shearer from being damaged or being involved in a safety accident.

#### IV. CONCLUSION

This research proposes a novel classifier selective ensemble method using the OSL-KELM classifier and neighborhood rough set attribute reduction technology to classify the height regulating pattern of the shearer drum. To improve the classification accuracy of the model, we introduce the migration attributes describing the shearing coal of the shearer. The multigranular attribute reduction technology based on the neighborhood rough sets is used to perturb the attributes space to generate different base classifiers. In addition, tenfold cross validation is used to tune hyperparameters of individual classifiers. The performance of the classifier ensemble on the field database is evaluated using statistical measures including accuracy, precision, recall and F1-score. The main results of this study are summarized as follows.

(1) The experimental results clearly demonstrate SDHRM-SEoKELMRS's superior performance in the height regulating pattern of the shearer drum, which the prediction accuracy is 84.68% and time requires 1.23s.

(2) The generalization performance of ensemble methods on data sets with migration features is higher than that on data sets without migration features, in which the accuracy has increased by up to 4.5%.

(3) The classification accuracy of SDHRM-SEoKELMRS which uses the proposed neighborhood rough reduction and the selective ensemble methods to remove effectively the irrelevant attributes and base classifiers has increased by 2.75% compared with the Random Subspace model.

(4) After comparing six base classifiers, the OSL-KELM method achieves the shortest time (1.23s) on the test set II, which has increased by 65.16% compared with the classifier SVM.

In summary, based on the observations above, SDHRM-SEoKELMRS shows great potential and advantages for shearer drum height regulating. In other words, we used this method for increasing prediction accuracy of the change law of the coal-rock interface and reducing classification error by selective ensemble. Moreover, the ensemble method can be easily extended to other classification problems in the fully mechanized coalface. It is worth mentioning that although the proposed method has advantages in calculating speed and accuracy, it still needs to be improved in order to meet the requirements of unmanned fully-mechanized mining coal.

Nevertheless, it is certainly possible that the complete industry data can improve the generalization ability of the ensemble model. So, future work should focus on enlarging the current database and investigating other combination methods to improve the generalization ability of the ensemble model.

#### REFERENCES

- [1] F. Liu, W. J. Cao, and J. M. Zhang, "Continuously promoting the coal mine intellectualization and the high-quality development of China's coal industry," *China Coal*, vol. 45, no. 12, pp. 32–37, 2019.

- [2] G. F. Wang, F. Liu, and Y. H. Pang, "Coal mine intellectualization: The core technology of high quality development," *J. China Coal Soc.*, vol. 44, no. 2, pp. 349–357, 2019.
- [3] S. B. Li, "Progress and development trend of intelligent mining technology," *Coal Sci. Technol.*, vol. 47, no. 10, pp. 102–110, 2019.
- [4] L. Yuan and P. S. Zhang, "Development status and prospect of geological guarantee technology for precise coal mining," *J. China Coal Soc.*, vol. 44, no. 8, pp. 2277–2284, 2019.
- [5] W. Li, C. M. Luo, H. Yang, and Q. G. Fan, "Memory cutting of adjacent coal seams based on a hidden Markov model," *Arabian J. Geosci.*, vol. 28, no. 6, pp. 15–25, 2013.
- [6] S. L. Bessinger and M. G. Nelson, "Remnant roof coal thickness measurement with passive gamma ray instruments in coal mines," *IEEE Trans. Ind. Appl.*, vol. 29, no. 3, pp. 562–565, May/Jun. 1993.
- [7] R. L. Chufo and W. J. Johnson, "A radar coal thickness sensor," *IEEE Trans. Ind. Appl.*, vol. 29, no. 5, pp. 834–840, Sep./Oct. 1993.
- [8] J. Ralston, D. Reid, C. Hargrave, and D. Hainsworth, "Sensing for advancing mining automation capability: A review of underground automation technology development," *Int. J. Mining Sci. Technol.*, vol. 24, no. 3, pp. 305–310, May 2014.
- [9] W. H. Chen, P. F. Nan, and X. H. Yan, "Prediction and model optimization of memory shearing trajectory based on deep learning," *J. China Coal Soc.*, vol. 2, no. 17, pp. 1–9, 2020.
- [10] W. Li, Q. G. Fan, and Y. Q. Wang, "An algorithm of shearer memory cutting based on grey-Markovian model," *J. Central South Univ., Sci. Technol.*, vol. 10, no. 42, pp. 3054–3058, 2011.
- [11] Q. Fan, W. Li, Y. Wang, L. Zhou, X. Yang, and G. Ye, "Control strategy for an intelligent shearer height adjusting system," *Mining Sci. Technol., China*, vol. 20, no. 6, pp. 908–912, Nov. 2010.
- [12] L. Si, Z. Wang, X. Liu, C. Tan, J. Xu, and K. Zheng, "Multi-sensor data fusion identification for shearer cutting conditions based on parallel quasi-Newton neural networks and the Dempster-Shafer theory," *Sensors*, vol. 15, no. 11, pp. 28772–28795, Nov. 2015.
- [13] Z. B. Wang, Z. P. Xu, and X. J. Dong, "Self-adaptive adjustment height of the drum in the shearer based on artificial immune and memory cutting," *J. China Coal Soc.*, vol. 34, no. 10, pp. 1405–1409, 2009.
- [14] T. M. Yang and S. B. Xiong, "Application of wavelet neural network for automatic ranging cutting height of shearer," *Wavelet Anal. Appl.*, vol. 12, no. 6, pp. 478–483, 2003.
- [15] H. Wang and Q. Zhang, "Dynamic identification of coal-rock interface based on adaptive weight optimization and multi-sensor information fusion," *Inf. Fusion*, vol. 51, pp. 114–128, Nov. 2019.
- [16] Z. P. Xu, Z. B. Wang, and J. P. Mi, "Shearer self-adaptive memory cutting," *J. Chongqing Univ.*, vol. 34, no. 4, pp. 134–140, 2011.
- [17] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
- [18] R. W. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," *Pattern Recognit. Lett.*, vol. 24, no. 6, pp. 833–849, Mar. 2003.
- [19] Q. Hu, D. Yu, and Z. Xie, "Neighborhood classifiers," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 866–876, Feb. 2008.
- [20] N. I. A. Wahab, A. Mohamed, and A. Hussain, "Feature selection and extraction methods for power systems transient stability assessment employing computational intelligence techniques," *Neural Process. Lett.*, vol. 35, no. 1, pp. 81–102, Feb. 2012.
- [21] H. H. Inbarani, M. Bagyamathi, and A. T. Azar, "A novel hybrid feature selection method based on rough set and improved harmony search," *Neural Comput. Appl.*, vol. 26, no. 8, pp. 1859–1880, Nov. 2015.
- [22] G. I. Webb and Z. Zheng, "Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 8, pp. 980–991, Aug. 2004.
- [23] H. Li, Y. Cui, Y. Liu, W. Li, Y. Shi, C. Fang, H. Li, T. Gao, L. Hu, and Y. Lu, "Ensemble learning for overall power conversion efficiency of the all-organic dye-sensitized solar cells," *IEEE Access*, vol. 6, pp. 34118–34126, 2018.
- [24] C.-F. Tsai, "Combining cluster analysis with classifier ensembles to predict financial distress," *Inf. Fusion*, vol. 16, no. 1, pp. 46–58, Mar. 2014.
- [25] J. Vijaya and E. Sivasankar, "An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing," *Cluster Comput.*, vol. 22, no. 5, pp. 1–12, 2017.
- [26] Q. Hu, D. Yu, Z. Xie, and X. Li, "EROS: Ensemble rough subspaces," *Pattern Recognit.*, vol. 40, no. 12, pp. 3728–3739, Dec. 2007.
- [27] I. Saha, J. Zubek, T. Klingström, S. Forsberg, J. Wikander, M. Kierczak, U. Maulik, and D. Plewczynski, "Ensemble learning prediction of protein-protein interactions using proteins functional annotations," *Mol. BioSyst.*, vol. 10, no. 4, pp. 820–830, 2014.
- [28] J. F. Zhang, Y. H. Wang, Y. T. Sun, and G. C. Li, "Strength of ensemble learning in multiclass classification of rockburst intensity," *Int. J. Numer. Anal. Methods Geomech.*, vol. 9, no. 1, pp. 1–21, 2020.
- [29] M. Mataric, "Environment learning using a distributed representation," in *Proc. IEEE Int. Conf. Robot. Automat.*, Cincinnati, OH, USA, May 1990, pp. 402–406.
- [30] C. Wang, Y. Huang, M. Shao, Q. Hu, and D. Chen, "Feature selection based on neighborhood self-information," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 4031–4042, Sep. 2020.
- [31] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines: A review," *Neural Netw.*, vol. 61, pp. 32–48, Jan. 2015.
- [32] J. Huang, Z. L. Yu, Z. Cai, Z. Gu, Z. Cai, W. Gao, S. Yu, and Q. Du, "Extreme learning machine with multi-scale local receptive fields for texture classification," *Multidimensional Syst. Signal Process.*, vol. 28, no. 3, pp. 995–1011, Jul. 2017.
- [33] L. Guo, J.-H. Hao, and M. Liu, "An incremental extreme learning machine for online sequential learning problems," *Neurocomputing*, vol. 128, pp. 50–58, Mar. 2014.
- [34] W. H. William, "Updating the inverse of a matrix," *SIAM Rev.*, vol. 32, no. 2, pp. 221–239, 1989.
- [35] M. Shahjahan, "A pruning algorithm for training cooperative neural network ensembles," *IEICE Trans. Inf. Syst.*, vol. E89-D, no. 3, pp. 1257–1269, Mar. 2006.
- [36] H. R. KoA, R. Sabourin, and A. S. Britto, "From dynamic classifier selection to dynamic ensemble selection," *Pattern Recognit.*, vol. 41, no. 5, pp. 173–1718, 2008.
- [37] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artif. Intell.*, vol. 137, nos. 1–2, pp. 239–263, May 2002.
- [38] G. Liu, Y. Xu, Z. He, Y. Rao, J. Xia, and L. Fan, "Deep learning-based channel prediction for edge computing networks toward intelligent connected vehicles," *IEEE Access*, vol. 7, pp. 114487–114495, 2019.
- [39] A. Zhukov, N. Tomin, V. Kurbatsky, D. Sidorov, D. Panasetsky, and A. Foley, "Ensemble methods of classification for power systems security assessment," *Appl. Comput. Informat.*, vol. 15, no. 1, pp. 45–53, Jan. 2019.



**ZHENGXIONG LU** received the B.Eng. degree in mechanical engineering from the Chongqing University of Science and Technology, in 2013, and the M.Eng. degree in mechanical engineering from the Xi'an University of Science and Technology, China, in 2016, where he is currently pursuing the Ph.D. degree with the School of Mechanical Engineering. His research interests include machine learning, information fusion, rough set, and their application in intelligent mining.



**WEI GUO** received the M.Eng. degree in mechanical engineering from the Xi'an University of Science and Technology, Xi'an, China, in 1988.

Since 1993, he has been a Professor with the Mechanical Engineering Department, Xi'an University of Science and Technology. Since 2012, he has been a Ph.D. Supervisor with the Xi'an University of Science and Technology. He has authored over 60 articles and conference proceedings in a well-known publishing platform. His

current research interests include the coal mine machinery design, mining equipment intelligent control, and intelligent mining technology. He was an excellent teacher in Shaanxi province. He received two provincial science and technology awards, the Municipal Science and Technology Award, and two provincial teaching achievement awards.



**CHUANWEI ZHANG** received the B.Eng. and M.Eng. degrees in mechatronic engineering from the Xi'an University of Science and Technology, Xi'an, China, in 2001, and the Ph.D. degree in mechatronic engineering from Xi'an Jiaotong University, Xi'an, in 2006.

From 2003 to 2007, he was a Lecturer with the Mechanical Engineering Department, Xi'an University of Science and Technology, where he has been a Professor, since 2007. He has authored three books, more than 50 articles, and more than 50 inventions. His research interests include intelligent network vehicles and autonomous navigation in complex environments.



**WUGANG ZHANG** received the Ph.D. degree from the Xi'an University of Science and Technology, Shaanxi, China, in July 2020. He is currently engaged in the research of mechanical engineering at the School of Mechanical Engineering, Xi'an University of Science and Technology. His main research interests include intellectualization of coal mine, laser tracking and positioning technology, and photoelectric detection technology. He has authored six articles on photoelectric

detection technology, tracking control method, machine learning algorithm, and so on.



**SHUANFENG ZHAO** received the B.Eng. degree in mechanical engineering from the China University of Mining and Technology, Xuzhou, China, in 2001, the M.Eng. degree in mechanical engineering from the Xi'an University of Science and Technology, Xi'an, China, in 2004, and the Ph.D. degree in instrument science and technology from Xi'an Jiaotong University, Xi'an, in 2013.

From 2006 to 2011, he was a Lecturer with the Mechanical Engineering Department, Xi'an University of Science and Technology, where he has been a Professor, Since 2011. His research interests include signal process, pattern recognition, deep learning, and monitor and control of mining equipment. He has published more than 20 articles in refereed journals and international conference proceedings in the above areas.



**MANZHI YANG** received the B.Eng. degree in agricultural mechanization and automation from the Shenyang Agricultural University, Shenyang, China, in 2009, and the M.A.Eng. degree and the Ph.D. degree in mechatronic engineering from the Xi'an University of Technology, Xi'an, in 2012 and 2016, respectively.

Since 2016, he has been a Lecturer with the Mechanical Engineering Department, Xi'an University of Science and Technology. He has authored ten articles and nine inventions. His research interests include precision mechanical and its motion control motion, and artificial intelligence.



**YUAN WANG** received the B.Eng. and M.A.Eng. degrees in mechanical manufacturing and automation from the Xi'an University of Science and Technology, Xi'an, in 2012, and the Ph.D. degree in mechanical engineering from the Xi'an University of Science and Technology, in 2016.

From 2012 to 2015, he was an Assistant Engineer with the Mechanical Engineering Department, Xi'an University of Science and Technology, where he has been an Engineer since 2015. He has authored three articles and three inventions. His research interests include coal mine automation control, machine vision, and image processing.



**SHUITIAN LI** received the B.Eng. and M.S. degrees in mechatronic engineering from the Xi'an University of Science and Technology, Xi'an, China, in 2014 and 2018, respectively.

Since 2018, he has been a Lecturer with the Mechanical Engineering Department, Xi'an University of Science and Technology. His research interest includes automatic drive.

...