

Received February 6, 2020, accepted February 17, 2020, date of publication February 27, 2020, date of current version March 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2976795

Features Fusion Exaction and KELM With Modified Grey Wolf Optimizer for Mixture Control Chart Patterns Recognition

MIN ZHANG^{1,2}, XUNJIE ZHANG¹, HAIBO WANG^{1,2},
GUOWEN XIONG¹, AND WENMING CHENG^{1,2}

¹School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China

²Technology and Equipment of Rail Transit Operation and Maintenance Key Laboratory of Sichuan Province, Chengdu 610031, China

Corresponding author: Haibo Wang (haibowang@swjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 51675450, and in part by the Ministry of Education in China (MOE) Project of Humanities and Social Sciences under Grant 18YJC630255.

ABSTRACT Control charts are significant diagnostic tools to detect and identify the quality fluctuation of the complex industrial process. In the practical production process, attention is being paid to the monitoring of mixture control charts, which usually coupled by two or more basic control charts modes. This research is to present a hybrid pattern recognition method for mixture control charts. The proposed method mainly covers the feature fusion extraction (FFE) and kernel extreme learning machine (KELM) with modified grey wolf optimizer (MGWO). The FFE module applies the original data and their shape and statistical features as the features, then uses kernel entropy component analysis to reduce the feature dimension and extract valid features. One significant difficulty of KELM is to get suitable parameters like the penalty parameter and the kernel function parameter value. MGWO is established to the optimal tuning of KELM parameters, which improves the population initialization and nonlinear convergence factor of traditional grey wolf optimizer. The proposed methodology is promising to obtain a better classification recognition rate, less computational time and achieves more stable results in the pattern recognition problem of mixture control charts.

INDEX TERMS Control chart patterns recognition, feature fusion extraction, kernel extreme learning machine, kernel entropy component analysis, modified grey wolf optimizer.

I. INTRODUCTION

With the development of data collection technology, control charts are the most popular tools in statistical process control and mainly apply to record or monitor the fluctuations in quality problems of complex industry. However, it is difficult to judge whether the control chart will lose control or not. Normal control chart pattern indicates that the process is under control. The abnormal patterns show that the process is out of control and there are some faults or variations needed to take the improvement actions [1]. As the production process becomes more complex and automated, the actual data present the multiple patterns coupling, which may be coupled by two or more basic modes. More details regarding the control chart patterns (CCPs) can be found, Figure 1 and Figure 2 show the six basic CCPs and four mixture CCPs, respectively [2]–[4].

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong.

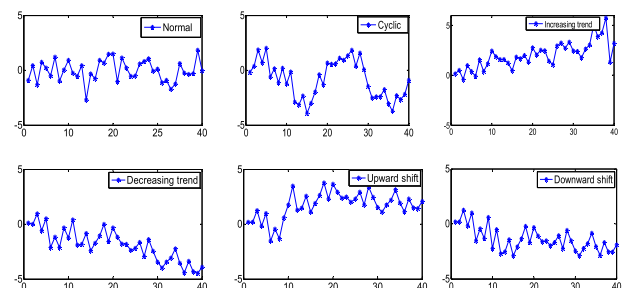


FIGURE 1. Six basic CCPs.

Adequate recognition of mixture CCPs is regarded as a crucial control risk to detect the fluctuations and explore the certain causes of abnormal patterns. Researchers have been developed various rules, such as elucidate manually, zone tests or expert systems, to help the engineers of quality control to recognize abnormal mixture CCPs [5]–[8]. Pelegrina *et al.* [5] proposed a blind source separation to recognize the concurrent CCPs. Lu *et al.* [6] utilized

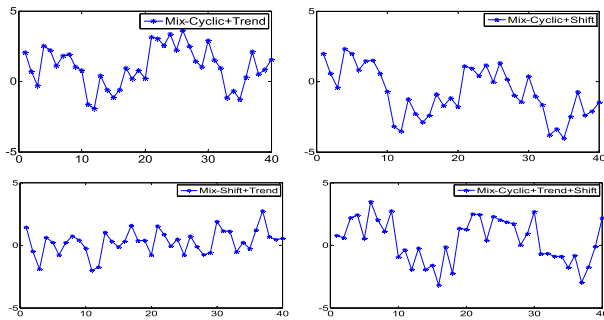


FIGURE 2. Four mixture CCPs.

independent component analysis to get efficient features, then recognized the mixture CCPs using support vector machine. Addeh *et al.* [7] described an optimized radial basis function neural network for CCPs. Yang *et al.* [8] identified mixture CCPs by extreme-point symmetric mode decomposition and extreme learning machine. These researchers offer some ways to realize mixture control charts pattern recognition more effective, however, it still has not addressed the problem adequately.

For the mixture CCPs problem, data processing plays an important role, in which the core principle is to transform the original data into feature data by dimensionality reduction [9], [10]. Multivariate statistical process monitoring methods, like partial least squares, principal component analysis (PCA), and some related improved methods like kernel PCA (KPCA), are the most classical dimensionality reduction methods [11]. However, these methods realize dimensionality reduction to describe the major trends by selecting the top eigenvalues of the process data set. Kernel entropy component analysis (KECA) is a novel spectral data processing approach, which reveals the intrinsic characteristics by the information entropy and not needs to select the eigenvectors of the kernel matrix. Moreover, transformed data of KECA have a distinct angular structure and each carries the cluster structure information, and it has been widely used to extract features [12]–[15]. However, data of most practical complex processes have the characteristics of time-varying, large amount and nonlinear. To solve these problems, a feature fusion extraction (FFE) method is proposed. The principle idea of this approach is to use KECA to obtain effective features from the original data and the shape and statistical features. This can decrease the features dimension and computational complexity.

Machine learning algorithms have been increasingly utilized in the classification of CCPs in recent years, like artificial neural networks (ANNs) and support vector machine (SVM). ANNs have been the most frequently used method, which aoolies a multi-layer perception with backpropagation to recognize CCPs [16]. However, ANNs still have many disadvantages, like the need for large amounts of data to train, poor generalization ability and over-fitting problem [9]. Recently, SVM has increased attention because it can obtain remarkable results. But its accuracy largely depends on the choice of the kernel function parameters [17], [18].

Extreme learning machine is an extremely efficient approach and has been widely applied in the financial, manufacturing and service industry [19], [20]. Kernel extreme learning machine (KELM) is proposed for the kernel matrix replaces the randomness matrix of ELM. The input data can be mapped to a high dimensional space by kernel functions, which can avoid the random fluctuations and find the hidden features more quickly and efficiently [21], [22]. The penalty factor and the kernel parameter play an important role in the output of KELM and should be optimized in this paper.

Presently, numerous studies have been shown biologically-inspired methods (like genetic algorithm [23], [24] and particle swarm optimization [25]) for the KELM kernel function parameter. However, these methods have the weaknesses of premature convergence and a relatively high percentage of errors. The grey wolf optimizer (GWO), which imitates the hunting behavior of grey wolves, has widely used in numerous studies on solving practical optimization problems [26]–[29]. Besides, many improved GWO methods have been proposed to enhance the searchability. Cai *et al.* [30] proposed a new parameter learning strategy based on an improved grey wolf optimization strategy. Zhao *et al.* [31] proposed chaos enhanced grey wolf optimization wrapped ELM for the diagnosis of paraquat-poisoned patients. Emary *et al.* [32] proposed a novel binary version of the grey wolf optimization to select the optimal feature subset for classification purposes. A modified grey wolf optimizer algorithm (MGWO), which improves the population initialization and nonlinear convergence factor, is applied to optimize the kernel function parameters of KELM.

In this paper, we propose many traditional approaches to verify the effectiveness of the proposed FFE_KELM_MGWO approach, like features, feature extraction methods, parameters optimization methods. Experimental results indicate that the proposed approach can effectively recognize mixture control charts patterns and perform better. The main specific findings of this paper can be summarized as follows:

1. A potential feature extraction approach FFE applies the original data and their shape and statistical features as the features, then uses KECA to reduce the feature dimension and extract valid features.
2. A new nature-inspired method MGWO, which improves the population initialization and nonlinear convergence factor, is applied to optimize the KELM kernel function parameters.
3. The proposed FFE_KELM_MGWO methodology tends to obtain better patterns recognition rate, less computational time and achieves more stable results in the mixture CCPs problem when compared to several other methods.

The remainder of this paper is organized as follows. Section 2 describes the needed concepts of the FFE method. In Section 3, the details of the KELM_MGWO classifier are elaborated. Section 4 introduces the detailed procedures of the FFE_KELM_MGWO methodology. In Section 5, the experimental results are given to verify the performance of

TABLE 1. Five shape features.

| Shape feature name | Formula | Interpretation |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
| Slope | $\beta_1 = \frac{\sum_{i=1}^m (t_i - \bar{t})(x_i - \bar{x})}{\sum_{i=1}^m (t_i - \bar{t})^2}$ where, $\bar{t} = \frac{\sum_{i=1}^m t_i}{m}$ | The slope of the least-square line $\bar{x} = \beta_1 t + \beta_0$, where, $\beta_0 = \bar{x} - \beta_1 \bar{t}$. |
| N_1 | For $i = 1, 2, \dots, m-1$ if $(x_i - \bar{x})(x_{i+k} - \bar{x}) < 0, n_{1i} = 1$ else $n_{1i} = 0$, and $N_1 = \sum_{i=1}^m n_{1i}$ | The number of mean crossing. |
| N_2 | For $i = 1, 2, \dots, m-1$ if $(x_i - \bar{x})(x_{i+k} - \bar{x}_{i+1}) < 0, n_{2i} = 1$ else $n_{2i} = 0$, where, $\bar{x}_i = \beta_1 t_i + \beta_0$, and $N_2 = \sum_{i=1}^m n_{2i}$ | The number of least-square line crossing. |
| APML | $APML = C_m^+ + C_m^-$ | The area between the pattern and its mean line. |
| APLS | $APLS = \sum_{i=1}^m x_i - \bar{x}_i $, where, for $i = 1, 2, \dots, m$ $\bar{x}_i = \beta_1 t_i + \beta_0$. | The area between the pattern and its least squares line. |

the proposed methods. Conclusions and the research findings are presented in Section 6.

II. THE PROPOSED FFE FEATURE EXTRACTION SCHEME

Complex production original data mostly express mixture control chart patterns. The original data usually present large amounts and coupling, extracting the suitable features is a challenging task. As we know, shape and statistical features can reflect the characteristics of the CCPs and have different properties of different types. To avoid the loss of important information, the original data, shape and statistical features are applied as the inputs, then use KECA to reduce the feature dimension and extract valid features.

A. SHAPE FEATURES

The shape features are efficient to simplify the data number and get useful information for mixture CCPs. Five sharp features are considered in this work including Slope, N_1 , N_2 , APML, APLS. The detail calculation method of these five shape features can be found in Table 1. Box plots of shape features in the different CCPs signals are shown in Figure 3.

B. STATISTICAL FEATURES

In Ref. [33], statistical features can obtain information for CCPs efficiently. We apply this method in the proposed method, which selects Mean, Standard deviation (SD), Mean-Square, Skewness, Kurtosis, Positive Cusum, Negative Cusum, and Average Autocorrelation as the parameters. Each statistical feature can be used to differentiate the patterns, their definitions are respectively shown in Table 2 and the values of these features for different CCPs signals are shown in Figure 4.

C. KECA FEATURE REDUCTION ALGORITHM

KECA is proposed by R. Jenssen, the original idea of the KECA method is introduced in detail [12], [34].

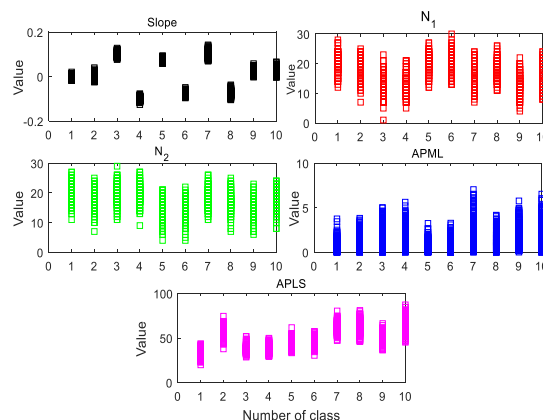


FIGURE 3. Box plots of the five shape features for different classes.

The definition of Renyi quadratic entropy is as follows:

$$H(p) = -\log \int p^2(x)dx \tag{1}$$

where $p(x)$ is a probability density function of the data set (X_1, X_2, \dots, X_n) . For the characteristics of the monotonic function, the quantity can be got by

$$V(p) = \int p^2(x)dx \tag{2}$$

A Parzen window density estimator is invoked to estimate $V(p)$ by

$$\hat{p}(x) = \frac{1}{N} \sum_{x_i \in D} k_\sigma(x, x_i) \tag{3}$$

where $k_\sigma(x, x_i)$ is a Mercer kernel function or Parzen window, σ is width parameter.

$$k_\sigma(x, x_i) = \exp(-\|x - x_i\|^2)/(2\sigma^2) \tag{4}$$

TABLE 2. Eight statistical features.

| Statistics feature name | Formula |
|--------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Mean (\bar{x}) | $\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i$ |
| Standard deviation(σ) | $\sigma = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2}$ |
| Mean-square value (\bar{x}^2) | $\bar{x}^2 = \frac{x_1^2 + x_2^2 + \dots + x_m^2}{m}$ $\bar{R}_{xx} = \frac{\sum_{k=0}^{m-1} R_{xx}(k)}{m}$, where |
| Average Autocorrelation (\bar{R}_{xx}) | $R_{xx}(k) = \frac{\sum_{i=1}^{m-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^m (x_i - \bar{x})^2}$ for $k=0, 1, \dots, m-1$ C_m^+ is calculated recursively by setting $C_0^+ = 0$ and for $1 \leq i \leq m$, $C_i^+ = \max[0, (x_i - \bar{x}) + C_{i-1}^+]$. C_i^+ is calculated recursively by setting $C_0^- = 0$ and for |
| Positive cusum (C_m^+) | $1 \leq i \leq m$, $C_i^- = \max[0, (\bar{x} - x_i) + C_{i-1}^-]$. |
| Negative cusum (C_m^-) | $1 \leq i \leq m$, $C_i^- = \max[0, (\bar{x} - x_i) + C_{i-1}^-]$. |
| Skewness (α_3) | $\alpha_3 = \frac{\sum_{i=1}^m (x_i - \bar{x})^3}{m\sigma^3}$ |
| Kurtosis (α_4) | $\alpha_4 = \frac{\sum_{i=1}^m (x_i - \bar{x})^4}{m\sigma^4}$ |

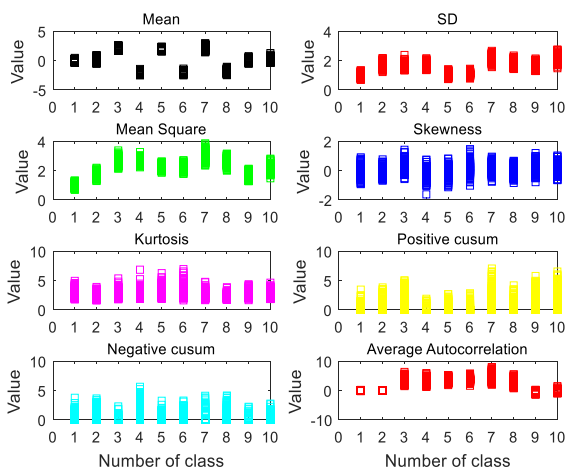


FIGURE 4. Box plots of the eight features for different classes.

Using the sample mean approximation of the expectation operator, we can obtain the following equation

$$\hat{V}(p) = \frac{1}{N} \sum_{x_i \in D} \hat{p}(x_i) = \frac{1}{N} \sum_{x_i \in D} \frac{1}{N} \sum_{x_{i'} \in D} k_{\sigma}(x_i, x_{i'}) \quad (5)$$

$$\hat{V}(p) = \frac{1}{N^2} \mathbf{I}^T \mathbf{K} \mathbf{I} \quad (6)$$

where element (t, t') of the $(N \times N)$ kernel matrix \mathbf{K} equals $k_{\sigma}(x, x_{t'})$ and \mathbf{I} is an $(N \times 1)$ vector of ones. \mathbf{K} is the kernel matrix:

$$\mathbf{K} = \mathbf{E} \mathbf{D} \mathbf{E}^T \quad (7)$$

where \mathbf{D} is a diagonal matrix of eigenvalues $\lambda_1, \dots, \lambda_n$; \mathbf{E} is a matrix of the corresponding eigenvectors e_1, e_2, \dots, e_n . Eq.(5) can be rewritten as

$$\hat{V}(p) = \frac{1}{N^2} \left(\sum_{i=1}^N \sqrt{\lambda_i} e_i^T \mathbf{I} \right)^2 \quad (8)$$

The transformation may be expressed as

$$\Phi_{eca} = \mathbf{P} \mathbf{U}_k \Phi = \mathbf{D}_k^{-1/2} \mathbf{E}_k^T : \min_{\lambda_1, e_1, \dots, \lambda_N, e_N} \hat{V}(p) - \hat{V}_k(p) \quad (9)$$

Out-of-sample data points represented by Φ' are projected onto \mathbf{U}_k over the selected components yielding.

$$\Phi'_{eca} = \mathbf{P} \mathbf{U}_k \Phi' = \mathbf{D}_k^{-1/2} \mathbf{E}_k^T \mathbf{K}'^T \quad (10)$$

In the dimension reduction of the KECA algorithm, the core is to estimate the contribution value of each principal element directly to the Rayleigh entropy, then determine the degree of retention of the information in the direction of the principal element, and then the data is mapped to kernel principal directions that contribute greatly to the Rayleigh entropy [14]. For retaining more information of original data, the entropy contribution rate is adopted to obtain the selected principal elements in the data dimensionality reduction process.

III. KELM_MGWO CLASSIFIER METHOD

A. KERNEL EXTREME LEARNING MACHINE (KELM)

Extreme learning machine (ELM) is proposed based on a single hidden layer feedforward neural network (SLFN), which has extremely generalization and fast learning ability compared to the traditional methods [19].

The advantage of ELM is that it can randomly generate the threshold of neurons in the hidden layer and the connection weight between the input and hidden layers. It does not need to adjust the weights of the SLFN, and hence involves less computational complexity and consumes less time.

The ELM structure is described below:

$$f_L(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x) \beta \quad (11)$$

where the output vector $\beta = [\beta_1, \beta_2, \dots, \beta_L]$, the ELM feature mapping function $h(x) = [h_1(x), h_2(x), \dots, h_L(x)]$; Input samples $x = [x_1, x_2, \dots, x_N]$, N is the number of patterns. It does not need to tune the initial weights, which is the input layer with m inputs and hidden layer with L neurons. Thus using the tanh function as the activation function:

$$h_i(x) = \tanh(w_{i0} + w_{i1}x_{i1} + w_{i2}x_{i2} + \dots + w_{im}x_{im}) \quad (12)$$

The hidden layer randomized matrix can be expressed as

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h(x) \cdots h(x_1) \\ \vdots \\ h(x_N) \cdots h_L(x_N) \end{bmatrix} \quad (13)$$

and the target vector is given by

$$T = \begin{bmatrix} t_1^* \\ t_2^* \\ \vdots \\ t_N^* \end{bmatrix} \quad (14)$$

where t^* means the expected output (sample label). Eq. (11) is expressed in a matrix form as

$$H\beta = T \quad (15)$$

To solve the overfitting problem and get better generalization ability, a constrained optimization method for β needs to establish in the original ELM. It can be expressed as

$$L_{CELM} = \frac{1}{2} \|\beta\|^2 + \frac{1}{2} C \|\xi\|^2 \quad (16)$$

$$\text{subject to } H\beta = T - \xi \quad (17)$$

where ξ is the error vector as $\xi = [\xi_1, \xi_2, \dots, \xi_N]$, C is the regularization parameter. KKT theorem is used to transform the constrained problem into a dual problem as

$$L_{DELM} = \frac{1}{2} \|\beta\|^2 + \frac{1}{2} C \|\xi\|^2 - \alpha (H\beta - T + \xi) \quad (18)$$

The output weight β is calculated as

$$\beta = \left(\frac{I}{C} + H^T H \right)^{-1} H^T T \text{ for } N > L \quad (19)$$

$$\text{or } \beta = H^T \left(\frac{I}{C} + H H^T \right)^{-1} T \text{ for } N < L \quad (20)$$

The output from the ELM is implemented as

$$f(x) = h(x) \left(\frac{I}{C} + H^T H \right)^{-1} H^T T \text{ for } N > L \quad (21)$$

$$\text{or } f(x) = h(x) H^T \left(\frac{I}{C} + H H^T \right)^{-1} T \text{ for } N < L \quad (22)$$

Compared with the traditional ANNs methods, the training process of ELM is simple. However, the learning performances, like stability and generalization, are determined by the selection of activation function and the neurons number in the hidden layer. To deal with this problem, KELM is designated using a kernel function to obtain better generalization and stability based on the Mercer theorem, the kernel matrix of the KELM is expressed as

$$K_{ELM} = H H^T, \text{ and } K_{ELM}(x_i, x_j) = h(x_i) h(x_j) \quad (23)$$

Hence, the output function is obtained as

$$f(x) = h(x) H^T (K_{ELM} + \frac{I}{C})^{-1} T \quad (24)$$

Eq. (24) can be written as follows:

$$f(x) = [(x, x_1), (x, x_2), \dots, (x, x_N)] (K_{ELM} + \frac{I}{C})^{-1} T \quad (25)$$

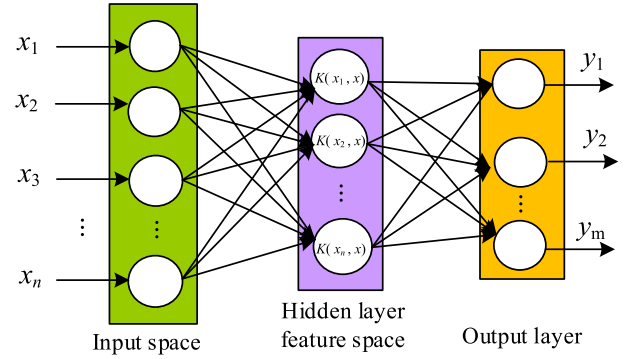


FIGURE 5. Box The structure of KELM model.

The hidden layer features of the maps can get by the kernel function, which does not need to know, the structure of the KELM model as shown in Figure 5. There are various kinds of kernel functions, radial basis function kernel (RBF) is the most widely adopted to solve non-linear problems in KELM. It can be defined as $K(x, x_i) = \exp(-\gamma \|x - x_i\|^2)$. Select the best values of the penalty parameter (C) and the kernel function parameter value (γ) are the crucial point of KELM. To overcome this problem, a modified grey wolf optimizer is applied to obtain the optimal parameter values.

B. MODIFIED GREY WOLF OPTIMIZER FOR TWO-DIMENSIONAL KELM PARAMETERS

The effectiveness of KELM is mainly determined by two crucial parameters of C and γ . In this paper, a modified grey wolf optimizer (MGWO) is proposed to optimize two-dimensional KELM parameters, where the good point set theory is used to generate the initialization of the gray wolf algorithm, and nonlinear hyperbolic tangent function as a new convergence is applied to balance the global searchability and the local development ability.

1) GREY WOLF OPTIMIZER

Grey wolf optimizer (GWO) is a newly introduced global optimization algorithm by Mirjalilli in 2014 [26], which imitates the hunting and searching mechanism of grey wolves.

The four-level social hierarchy of grey wolves in GWO has assumed α , β , δ and ω . α wolves are considered as the leaders and all other grey wolves follow their instructions. β wolves are responsible for helping α in their decision making, which is the best candidate to be the α . δ wolves dominate the wolves of the last level ω wolves. The detail social hierarchy is shown in Figure 6.

The distances from α , β and δ wolves i.e. D_α , D_β and D_δ to each of the remaining wolf (\vec{X}) are obtained using Eq. (26). \vec{X}_1 , \vec{X}_2 and \vec{X}_3 can be obtained as expressed in Eq. (27).

$$\vec{D}_\alpha = \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right|, \vec{D}_\beta = \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right|, \quad (26)$$

$$\vec{D}_\delta = \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right|$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha, \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta, \quad (27)$$

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta$$

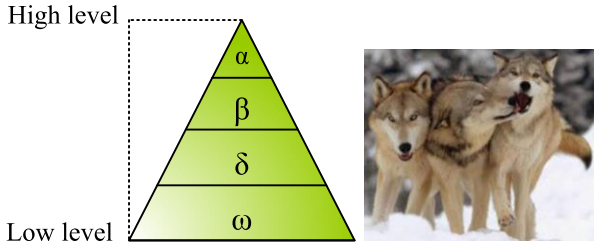


FIGURE 6. The social hierarchy of grey wolves.

$$\vec{A}_2 = 2\vec{a} \cdot \vec{r}_1 - \vec{a}, \vec{C} = 2 \cdot \vec{r}_2 \quad (28)$$

$$\vec{X}(t + 1) = (\vec{X}_1 + \vec{X}_2 + \vec{X}_3)/3 \quad (29)$$

where α , A and C are calculated using Eq. (28), \vec{r}_1 and \vec{r}_2 are the random numbers between $[0, 1]$. Vector \vec{a} is used in calculating \vec{A} , which involves the controlling activity in GWO. The element of \vec{a} linearly decreases from 2 to 0 at each process of iteration. \vec{C} can set some extra weight on the prey, it makes the wolves difficult to find it. Finally, their positions $\vec{X}(t + 1)$ of other wolves update using Eq. (29).

2) MODIFIED GREY WOLF OPTIMIZER

To get the better solution of two-dimensional KELM parameters, a modified grey wolf optimizer is proposed, where the good point set theory is used to generate the initialization of the gray wolf algorithm and nonlinear hyperbolic tangent function as a new convergence is applied. A detailed description of the modified grey wolf optimizer is introduced in this section.

a: IMPROVEMENT OF POPULATION INITIALIZATION

The diversity of the initial population affects the global convergence speed and the quality of the solution of the swarm intelligence algorithm, better diversity can improve the optimization performance. In the case of the global optimal solution unknown, the initial population is spread within the search range as widely as possible, which can provide superior global search performance in the early stage of the algorithm [23], [35].

The standard gray wolf optimization algorithm uses a random method to generate the initial population, which is difficult to guarantee the diversity of the initial population. Besides, it also makes the algorithm poorly diverse and affects the global search ability [24]. To improve search efficiency and ensure the diversity of the initial population, this paper uses a great point set theory to generate the initial population. This method makes the initial population more evenly distributed in the solution space.

According to the theory of good point set, the method of making good points in t -dimensional space is as follows:

$$P_n(k) = \{(\{r_1 * k\}, \dots, \{r_i * k\}, \dots, \{r_t * k\}), k = 1, 2, \dots, m\} \quad (30)$$

where $r_i = \{2 \cos(2\pi i/p)\}$, $1 \leq i \leq t$, p is the smallest prime number, which is satisfied $p \geq 2t + 3$. In this paper,

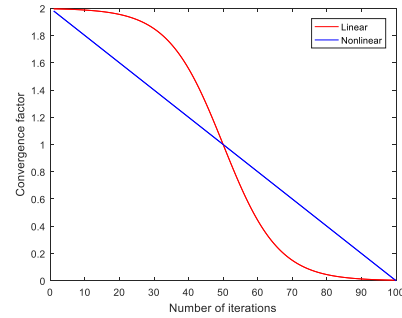


FIGURE 7. Decreasing factor curve.

the solution space dimension $t = 2$, that is $p \geq 7$, that is, p is the smallest prime number greater than or equal to 7, and the specific value of p is 7. $r_i = e^t$, $1 \leq i \leq t$, $\{r_i * k\}$ is the fractional part of $r_i * k$.

b: IMPROVEMENT OF NONLINEAR CONVERGENCE FACTOR

The convergence factor a in the standard gray wolf algorithm is a very important parameter, which decreases linearly from 2 to 0. It can adjust the balance of global and local search, where the large initial value a of the algorithm guarantees the global searchability, the later smaller value a enhance the searchability near the current optimal solution and ensure the local development capability. In the complex algorithm search process, the linear convergence factor can not fully balance the global search ability and local development ability [25].

For improving the recognition accuracy and search speed, this paper proposes a new convergence factor update formula based on hyperbolic tangent function. The update convergence factor a equation is:

$$a(t) = \tanh(\lambda \frac{t}{t_{\max}} + k) + \frac{a_{\text{initial}} - a_{\text{final}}}{2} \quad (31)$$

where t_{\max} is the maximum number of iterations, this paper is set as 100; a_{initial} and a_{final} are the initial and final values of the convergence factor a , are 2 and 0, respectively; λ and k are the adjustment parameters, λ is the deceleration rate of the adjustment factor, k is the adjustment factor, which can further adjust the global search and local development capabilities, $\lambda = -2\pi$, $k = \pi$. According to the parameter settings in this paper, the decreasing curve of the convergence factor a is shown in Figure 7.

The improved nonlinear decrement factor can balance the global search ability and local development ability. A large value is maintained for a long time in the early stage to improve global search efficiency. In the latter part of the iteration, it keeps a small value for a long time to improve the local development and the local search accuracy.

IV. THE PROPOSED FFE_KELM_MGWO SCHEME

This study integrates FFE, KELM, and MGWO to effectively and automatically recognize mixture CCPs. The proposed scheme is divided into three parts: feature extraction, parameter optimization, and pattern recognition. The general

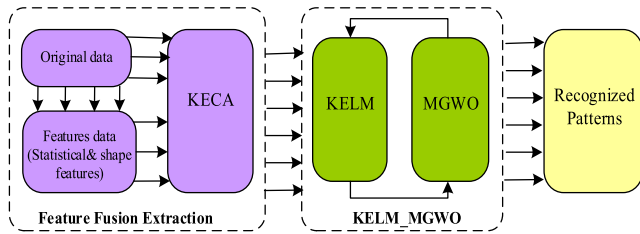


FIGURE 8. The general structure of the proposed method.

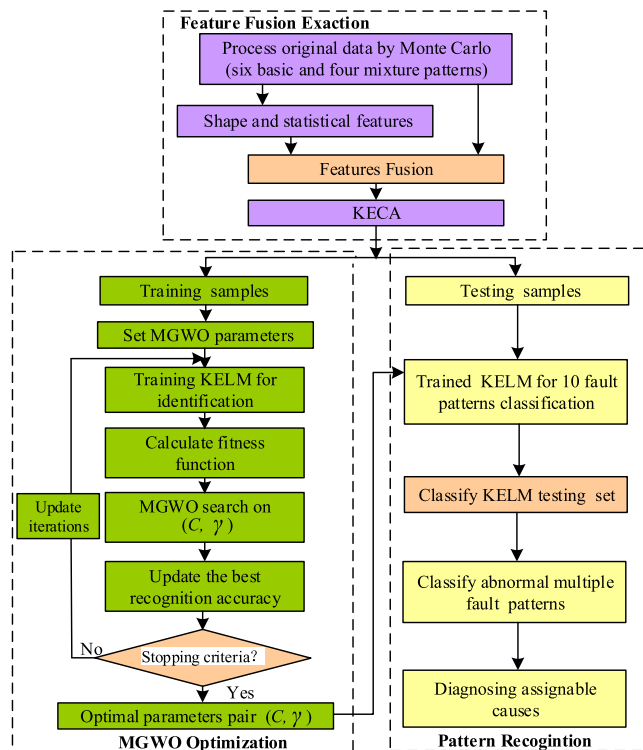


FIGURE 9. Flow chart of FFE_KELM_MGWO model.

structure and the detailed schematic diagram can be described in Figure 8 and Figure 9.

In the feature extraction part, FFE is utilized to obtain effective features. The original data of the process usually present large amounts and coupling, extracting the suitable features is a challenging task. As we know, the shape and statistical features of CCPs are different and can be applied to distinguish the mixture CCPs. However, it has the problem of difficulty to represent the complete information. To overcome this matter, the original data and the shape and statistical features are mixed as the features input. However, it makes the feature dimension increases sharply, which increases the computational difficulty. The KECA method, which determines the selected principal elements by the entropy contribution rate, is utilized to decrease the feature dimension.

In the parameters optimization part, MGWO is employed to optimize two key parameters of KELM classifier: the penalty parameter (C) and the kernel function parameter (γ). The classification accuracy of training samples is applied as

the fitness function:

$$F = ACC = \frac{k}{N} \quad (32)$$

where let N be the total number of training samples, k is the number of correctly identified labels, and ACC is the accuracy of the trained model by KEML classifier. The average accuracy is obtained as the ratio of correctly identified labels to the trained labels. Optimal parameters pair (C , γ) of KELM get by modified grey wolf optimizer.

V. PERFORMANCE ANALYSES

The performances of the proposed approach are evaluated in this section. The mixture CCPs are obtained from an automatic pattern generator, the formulas presented in Table 3 [30]. Six basic, normal (NOR), cyclic (CYC), increasing trend (IT), decreasing trend (DT), upward shift (US) and downward shift (DS), are considered. Four mixture CCPs are coupled by two or more basic modes. The principle of upward/downward shift or increasing/decreasing trend is similar, one of them is separately selected in this study. The increasing trend, downward shift and cyclic are applied to simulate the mixture CCPs.

For this study, each pattern contains 200 samples, which the observation window used is 40 data points. We have used 70% of the data for training and the rest for testing. The performance is evaluated by the average recognition accuracy of the mixtures CCPs on the testing samples. The average accuracy is obtained as the ratio of correctly recognized labels to the tested labels. Computer programs of experiments are performed on MATLAB [R2016b] environment in Intel core i3-4150 with 4 GB RAM using Dell computer.

A. PERFORMANCE OF FFE METHOD EVALUATION

Feature extraction plays an important part in the effectiveness and speed of mixture CCPs. To explore the superior performance of the proposed FFE approach, two other methods are used to compare: original data (OD) and shape and statistical features (SSF). OD method chooses the 40-dimensional time-domain sample data as features. SSF uses 8 statistical features and 5 shape features as the features. KELM with MGWO optimization algorithm is utilized to identify the mixture CCPs.

In this paper, the performances of the three feature extraction methods are evaluated using the recognition accuracy rate (RAR) and run time. The minimum (Min), mean, and the maximum (Max) of RAR, standard deviation, optimal parameters best (C , γ) and run time are calculated from the 10 independent runs, the RAR results are shown in Table 4.

From Table 4 and Figure 10, it can be seen that the average recognition accuracy of the proposed FFE_KELM model has arrived at 98.66%, which is better recognition performance. Comparing the standard deviation of three methods, FFE feature extraction method is significantly reduced. It indicates that the proposed feature extraction approach can achieve global optimal solutions more stable. Besides, it shows that the three methods can finish within a suitable run time, FFE has not increased the computational burden.

TABLE 3. The parameters of mixture CCPs.

| NO. | type | patterns | Generator equations | Parameters |
|-----|-----------|------------------------------------------------|--------------------------------------------------------------------|----------------------------------------------------------------|
| 1 | Basic 1 | Normal(NOR) | $X(t)=\mu+r\times\sigma$ | Mean $\mu=0$, |
| 2 | Basic 2 | Cyclic (CYC) | $X(t)=\mu+r\times\sigma+asin(2\pi t/t')$ | SD $\sigma=1$ |
| 3/4 | Basic 3/4 | Increasing/Decreasing Trend (IT/DT) | $X(t)=\mu+r\times\sigma\pm g\times t$ | Period $t'=(8\sigma$ to |
| 5/6 | Basic 5/6 | Increasing/Decreasing Shift (US/DS) | $X(t)=\mu+r\times\sigma\pm k\times s(k=1$ if $t > P$, else $k=0)$ | $16\sigma)$, amplitude $\alpha=(1.5\sigma$ to $2.5\sigma)$ |
| 7 | Mix1 | Cyclic + Increasing Trend (CT) | $X(t)=\mu+r\times\sigma+asin(2\pi t/t') + g\times t$ | Gradient $g=(0.05\sigma$ to $0.15\sigma)$ |
| 8 | Mix2 | Cyclic + Decreasing Shift (CS) | $X(t)=\mu+r\times\sigma+asin(2\pi t/t')-k\times s$ | Shift magnitude |
| 9 | Mix3 | Increasing Trend + Decreasing Shift (TS) | $X(t)=\mu+r\times\sigma+g\times t -k\times s$ | $s=(1.5\sigma$ to $2.5\sigma)$, |
| 10 | Mix4 | Cyclic+Increasing Trend+Decreasing Shift (CTS) | $X(t)=\mu+r\times\sigma+asin(2\pi t/t') + g\times t -k\times s$ | shift position $P=2$ |

TABLE 4. Recognition accuracy rate of three feature extraction methods.

| Feature extraction method | Best (C, γ) | RAR (/%) | | | Standard deviation | Run time (/s) |
|--------------------------------------|---------------------|--------------|--------------|--------------|--------------------|-----------------|
| | | Min | Mean | Max | | |
| Original data (OD) | (9.04, 8.74) | 65.83 | 88.38 | 98.50 | 14.28 | 18.45 |
| Shape and statistical features (SSF) | (590.9, 84.6) | 76.67 | 90.45 | 96.33 | 5.81 | 19.12 |
| Fusion Feature Reduction(FFE) | (98.9, 653.0) | 97.17 | 98.66 | 99.50 | 0.67 | 18.56 |

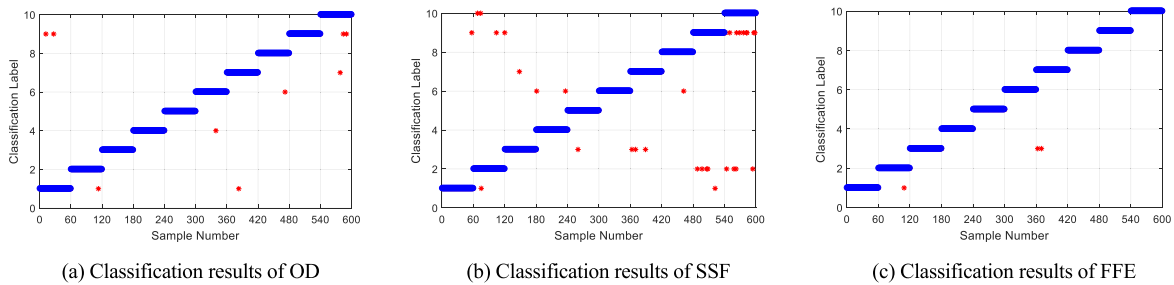


FIGURE 10. RAR results of the three methods: (a) OD, (b) SSF, (c) FFE.

B. PERFORMANCE OF RECOGNIZER IN THE STATISTICAL PROCESS CONTROL METHOD

The importance of the optimal selection of feature extraction is proved. The KECA method is evaluated in this experiment, which is used to get the feature dimensionality reduction for mixture CCPs classifier. We apply the traditional statistical process control methods like PCA and KPCA. KELM with MGWO optimization approach is used to recognize the mixture CCPs. The average recognition accuracies and run time of the three statistical process control models are presented in Table 5.

Table 5 illustrates that 99.50% recognition accuracy rate is achieved by FFE in the KECA method, 89.62% and 98.67% are got under the FFE in PCA, KPCA method. The proposed KECA statistical process control model has the best recognition accuracy rate. Besides, by comparing with OD and SSF, the FFE method can obtain better RAR in different statistical process control methods, it indicates that FFE is valid once again. Table 5 lists the run time of the proposed

TABLE 5. Comparison of three statistical process control methods.

| Method | Feature extraction method | Recognition Accuracy Rate (%) | Run time (/s) |
|--------|---------------------------|-------------------------------|---------------|
| PCA | OD | 78.88 | 38.79 |
| | SSF | 85.17 | 28.42 |
| | FFE | 89.62 | 29.62 |
| KPCA | OD | 94.33 | 35.08 |
| | SSF | 93.33 | 34.87 |
| | FFE | 98.67 | 35.06 |
| KECA | OD | 98.50 | 18.45 |
| | SSF | 94.67 | 18.57 |
| | FFE | 99.50 | 18.56 |

KECA statistical process control method mostly less than the other two methods. FFE_KELM_MGWO using the KECA method can obtain the best recognition accuracy rate and least run time.

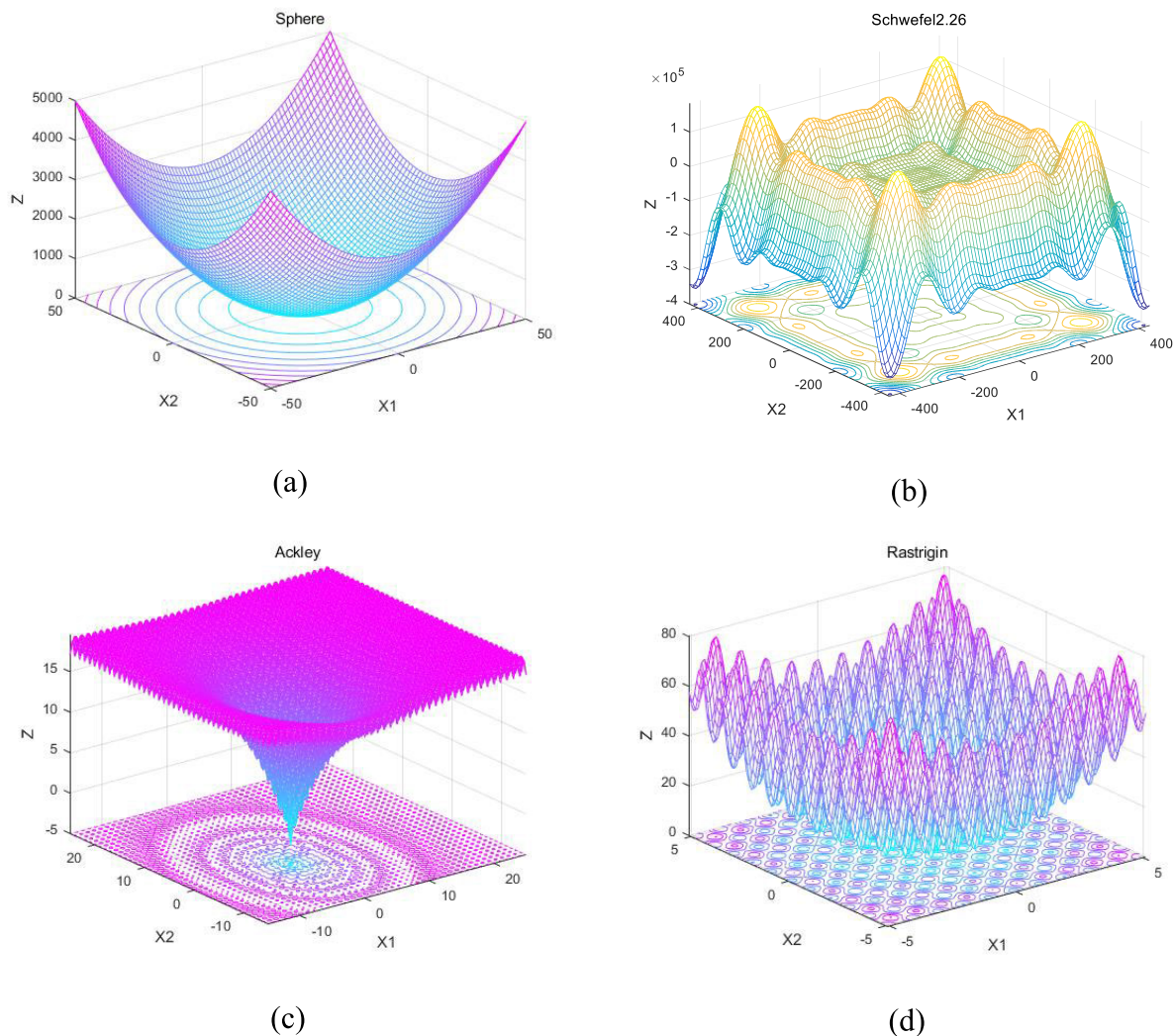


FIGURE 11. Four benchmark functions: (a) Sphere, (b) Schwefel2.26, (c) Ackley, and (d) Rastrigin.

C. PERFORMANCE OF MGWO OPTIMIZATION METHOD

To illustrate the effectiveness of the proposed MGWO method, the following four benchmark functions are used for verification. They are Sphere, Schwefel2.26, Ackley, and Rastrigin functions, respectively. The figure 11 shows the functions images.

To investigate the effectiveness of KELM_MGWO, three other methods GA, PSO, and GWO are utilized. The convergence results of the above four benchmark functions are shown as figure 12, it can be seen that the proposed MGWO method in this paper has great performance, fast convergence speed, and superior global search capability.

The performance of MGWO parameter optimization algorithms of KELM is evaluated for mixture CCPs. Two key parameters of the best penalty parameter (C) and the kernel function parameter (γ) of KELM should be chosen in the experiment. To investigate the effectiveness of KELM_MGWO, three other methods GA, PSO, and GWO are utilized. The results of the average recognition accuracies and run time of the four optimization models are described in Table 6.

TABLE 6. Comparison of three parameters optimization methods.

| Method | Feature extraction method | RAR (%) | Run time (s) |
|-----------|---------------------------|--------------|---------------|
| KELM_GA | OD | 98.33 | 585.78 |
| | SSF | 94.5 | 595.99 |
| | FFE | 99.33 | 635.58 |
| KELM_PSO | OD | 98.33 | 403.69 |
| | SSF | 94.67 | 386.17 |
| | FFE | 99.33 | 377.79 |
| KELM_GWO | OD | 98.50 | 17.80 |
| | SSF | 94.50 | 18.15 |
| | FFE | 99.33 | 18.11 |
| KELM_MGWO | OD | 98.50 | 18.45 |
| | SSF | 94.67 | 18.57 |
| | FFE | 99.50 | 18.56 |

Table 6 illustrates the best recognition accuracies in FFE of the KELM_GA, KELM_PSO, and KELM_GWO are the same as 99.33%, respectively. It indicates that

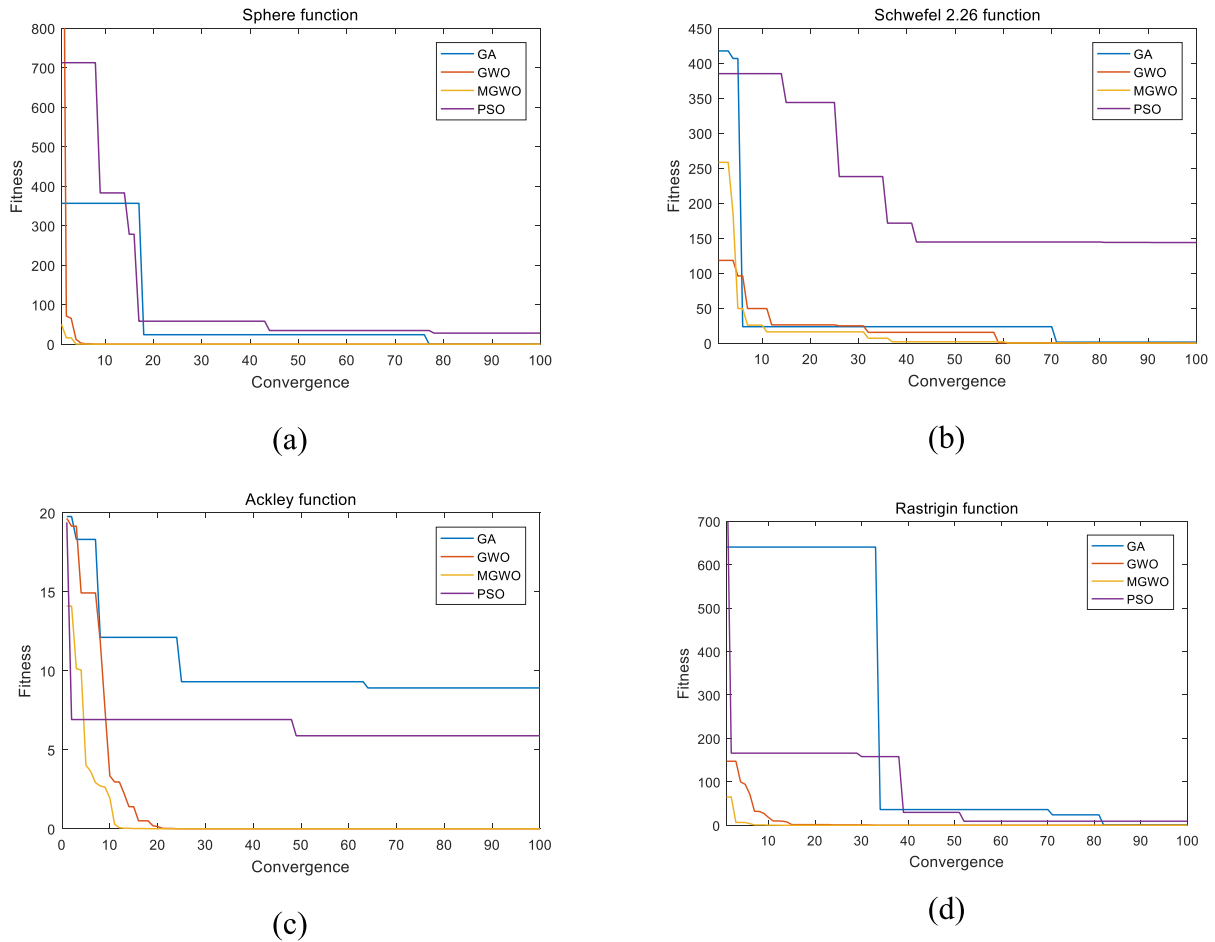


FIGURE 12. The results of four benchmark functions in four different methods: (a) Sphere, (b) Schwefel 2.26, (c) Ackley, and (d) Rastrigin.

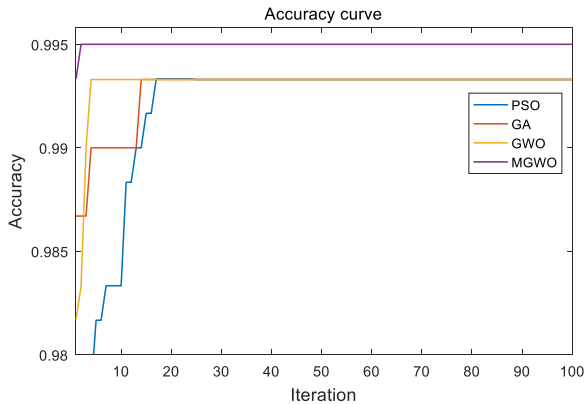


FIGURE 13. The convergence curves of four different optimization methods.

optimization methods of KELM can get a relatively good solution in recognition accuracy under FFE. The proposed FFE_KELM_MGWO model can obtain at 99.50%, which has a better recognition accuracy rate. At the same time, figure 13 shows the convergence curves of four different optimization methods. It indicates that the proposed method can obtain the optimal parameters.

In different parameter optimization methods, the FFE method always obtains the better RAR. However, the run

time of the proposed MGWO and GWO parameters optimization methods are significantly reduced compared with GA and PSO. It indicates that the GWO method can improve computational efficiency.

VI. CONCLUSION

Abnormal control chart patterns can reflect the quality fluctuation of the complex industrial process. This work identifies problems by mixture CCPs, in which the characteristics are coupled by two or more basic modes. A fast intelligent and accurate approach for mixture CCPs is proposed in the study, which consists of two main aspects: feature fusion exaction and parameter optimization in the KELM classifier. Original data, shape and statistical features, features fusion and KECA are utilized to make the features more effective. Modified grey wolf optimizer is next utilized to obtain the best two-dimensional KELM parameters.

In the experiments, the basic and mixtures patterns are simulated by the works of literature. From the computational results, the proposed FFE_KELM_MGWO is quite effective in identifying the mixture CCPs. Our main contributions can be summarized in three aspects. Firstly, FFE is successfully applied to combine the original data and the shape and statistical features, then uses KECA to reduce the feature dimension and extract valid features. Secondly, the experiment also

shows that the proposed KECA statistical process control method can deliver better recognition results compared with PCA and KPCA. Third, the proposed MGWO algorithm in training the KELM classifier can get a better combination of parameters, both the recognition accuracy rate and run time are greatly improved.

Currently, the proposed scheme is quite effective in recognizing mixture control charts. An enhancement would be to analyze the specific failure causes and predict fault in a few time steps in advance. More works will be contributed to promoting the widespread use of the proposed algorithm in the actual complex production process.

REFERENCES

- [1] K. W. Khaw, M. B. C. Khoo, P. Castagliola, and M. A. Rahim, "New adaptive control charts for monitoring the multivariate coefficient of variation," *Comput. Ind. Eng.*, vol. 126, pp. 595–610, Dec. 2018.
- [2] V. Ranaee, A. Ebrahimzadeh, and R. Ghaderi, "Application of the PSO-SVM model for recognition of control chart patterns," *ISA Trans.*, vol. 49, no. 4, pp. 577–586, Oct. 2010.
- [3] Y. E. Shao and C.-C. Chiu, "Applying emerging soft computing approaches to control chart pattern recognition for an SPC-EPC process," *Neurocomputing*, vol. 201, pp. 19–28, Aug. 2016.
- [4] N. Chukhrova and A. Johannssen, "Improved control charts for fraction non-conforming based on hypergeometric distribution," *Comput. Ind. Eng.*, vol. 128, pp. 795–806, Feb. 2019.
- [5] G. D. Pelegrina, L. T. Duarte, and C. Jutten, "Blind source separation and feature extraction in concurrent control charts pattern recognition: Novel analyses and a comparison of different methods," *Comput. Ind. Eng.*, vol. 92, pp. 105–114, Feb. 2016.
- [6] C.-J. Lu, Y. E. Shao, and P.-H. Li, "Mixture control chart patterns recognition using independent component analysis and support vector machine," *Neurocomputing*, vol. 74, no. 11, pp. 1908–1914, May 2011.
- [7] A. Addeh, A. Khormali, and N. A. Golilarz, "Control chart pattern recognition using RBF neural network with new training algorithm and practical features," *ISA Trans.*, vol. 79, pp. 202–216, Aug. 2018.
- [8] W.-A. Yang, W. Zhou, W. Liao, and Y. Guo, "Identification and quantification of concurrent control chart patterns using extreme-point symmetric mode decomposition and extreme learning machines," *Neurocomputing*, vol. 147, pp. 260–270, Jan. 2015.
- [9] M. Zhang and W. Cheng, "Recognition of mixture control chart pattern using multiclass support vector machine and genetic algorithm based on statistical and shape features," *Math. Problems Eng.*, vol. 2015, pp. 1–10, 2015.
- [10] S. X. Ding, "Data-driven design of monitoring and diagnosis systems for dynamic processes: A review of subspace technique based schemes and some recent results," *J. Process Control*, vol. 24, no. 2, pp. 431–449, Feb. 2014.
- [11] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annu. Rev. Control*, vol. 36, no. 2, pp. 220–234, Dec. 2012.
- [12] R. Jenssen, "Kernel entropy component analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 847–860, May 2010.
- [13] J. Shi, Q. Jiang, R. Mao, M. Lu, and T. Wang, "FR-KECA: Fuzzy robust kernel entropy component analysis," *Neurocomputing*, vol. 149, pp. 1415–1423, Feb. 2015.
- [14] H. Zhang, Y. Qi, L. Wang, X. Gao, and X. Wang, "Fault detection and diagnosis of chemical process using enhanced KECA," *Chemometric Intell. Lab. Syst.*, vol. 161, pp. 61–69, Feb. 2017.
- [15] J. Shi, Q. Jiang, Q. Zhang, Q. Huang, and X. Li, "Sparse kernel entropy component analysis for dimensionality reduction of biomedical data," *Neurocomputing*, vol. 168, pp. 930–940, Nov. 2015.
- [16] H. De la Torre Gutierrez and D. T. Pham, "Estimation and generation of training patterns for control chart pattern recognition," *Comput. Ind. Eng.*, vol. 95, pp. 72–82, May 2016.
- [17] A. Khormali and J. Addeh, "A novel approach for recognition of control chart patterns: Type-2 fuzzy clustering optimized support vector machine," *ISA Trans.*, vol. 63, pp. 256–264, Jul. 2016.
- [18] C. Wu, F. Liu, and B. Zhu, "Control chart pattern recognition using an integrated model based on binary-tree support vector machine," *Int. J. Prod. Res.*, vol. 53, no. 7, pp. 2026–2040, Sep. 2014.
- [19] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2004, pp. 985–990.
- [20] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1033–1044, May 2014.
- [21] P. P. Das, R. Bisoi, and P. K. Dash, "Data decomposition based fast reduced kernel extreme learning machine for currency exchange rate forecasting and trend analysis," *Expert Syst. Appl.*, vol. 96, pp. 427–449, Apr. 2018.
- [22] S. Lu, H. Yu, H. Dong, X. Wang, and Y. Sun, "Single-step prediction method of burning zone temperature based on real-time wavelet filtering and KELM," *Eng. Appl. Artif. Intell.*, vol. 70, pp. 142–148, Apr. 2018.
- [23] F. Ertam and E. Avci, "A new approach for Internet traffic classification: GA-WK-ELM," *Measurement*, vol. 95, pp. 135–142, Jan. 2017.
- [24] B. Liu, L. Tang, J. Wang, A. Li, and Y. Hao, "2-D defect profile reconstruction from ultrasonic guided wave signals based on QGA-kernelized ELM," *Neurocomputing*, vol. 128, pp. 217–223, Mar. 2014.
- [25] L. Zhang and J. Yuan, "Fault diagnosis of power transformers using kernel based extreme learning machine with particle swarm optimization," *Appl. Math. Inf. Sci.*, vol. 9, no. 2, pp. 1003–1010, 2015.
- [26] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.
- [27] M. Kohli and S. Arora, "Chaotic grey wolf optimization algorithm for constrained optimization problems," *J. Comput. Des. Eng.*, vol. 5, no. 4, pp. 458–472, Oct. 2018.
- [28] M. A. Al-Betar, M. A. Awadallah, H. Faris, I. Aljarah, and A. I. Hammouri, "Natural selection methods for grey wolf optimizer," *Expert Syst. Appl.*, vol. 113, pp. 481–498, Dec. 2018.
- [29] M. Wang, H. Chen, H. Li, Z. Cai, X. Zhao, C. Tong, J. Li, and X. Xu, "Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction," *Eng. Appl. Artif. Intell.*, vol. 63, pp. 54–68, Aug. 2017.
- [30] Z. Cai, J. Gu, J. Luo, Q. Zhang, H. Chen, Z. Pan, Y. Li, and C. Li, "Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy," *Expert Syst. Appl.*, vol. 138, Dec. 2019, Art. no. 112814.
- [31] X. Zhao, X. Zhang, Z. Cai, X. Tian, X. Wang, Y. Huang, H. Chen, and L. Hu, "Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients," *Comput. Biol. Chem.*, vol. 78, pp. 481–490, Feb. 2019.
- [32] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016.
- [33] M. Zhang, Y. Yuan, R. Wang, and W. Cheng, "Recognition of mixture control chart patterns based on fusion feature reduction and fire-works algorithm-optimized MSVM," *Pattern Anal. Appl.*, vol. 23, no. 1, pp. 15–26, 2020.
- [34] Q. Jiang, X. Yan, Z. Lv, and M. Guo, "Fault detection in nonlinear chemical processes based on kernel entropy component analysis and angular structure," *Korean J. Chem. Eng.*, vol. 30, no. 6, pp. 1181–1186, Apr. 2013.
- [35] W.-Y. Deng, Y.-S. Ong, and Q.-H. Zheng, "A fast reduced kernel extreme learning machine," *Neural Netw.*, vol. 76, pp. 29–38, Apr. 2016.



MIN ZHANG received the B.S. and Ph.D. degrees from the School of Mechanical Engineering, Southwest Jiaotong University, Chengdu, China, in 2004, and 2013, respectively.

She is currently a Lecturer (equivalent to an Assistant Professor) with the School of Mechanical Engineering and the Technology and Equipment of Rail Transit Operation and Maintenance Key Laboratory of Sichuan Province, Southwest Jiaotong University. Her research inter-

ests include intelligent fault diagnosis, quality control, machine learning, and optimization.



XUNJIE ZHANG received the B.S. degree from the School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing, China, in 2019. He is currently pursuing the M.S. degree with the School of Mechanical Engineering, Southwest Jiaotong University, Chengdu, China. His research interest is focused on intelligent fault diagnosis.



GUOWEN XIONG received the B.S. degree from the Renai College of Tianjin University, Tianjin, china, in 2017. He is currently pursuing the M.S. degree in mechanical engineering, Southwest Jiaotong University, Chengdu. His research interests include vehicle route optimization and process optimization.



HAIBO WANG received the B.S. and Ph.D. degrees from the School of Mechanical and Energy Engineering, Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively.

He is currently an Associate Professor with the School of Mechanical Engineering and the Technology and Equipment of Rail Transit Operation and Maintenance Key Laboratory of Sichuan Province, Southwest Jiaotong University. His research interests include intelligent control systems, pattern recognition, and intelligent fault diagnosis.



WENMING CHENG received the B.S., M.S., and Ph.D. degrees from the School of Mechanical Engineering, Southwest Jiaotong University, Chengdu, China, in 1988, 1989, and 2000, respectively.

He is currently a Professor with the School of Mechanical Engineering and the Technology and Equipment of Rail Transit Operation and Maintenance Key Laboratory of Sichuan Province, Southwest Jiaotong University. His research interests include mechanical intelligent optimization and dynamic simulation, logistics technology and equipment, and intelligent fault diagnosis.

...