

Received March 15, 2021, accepted March 23, 2021, date of publication March 26, 2021, date of current version April 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3068959

An Enhanced Unsupervised Extreme Learning Machine Based Method for the Nonlinear Fault Detection

LANYUN SHAO¹, RONGBAO KANG², WEILIN YI³, AND HANYUAN ZHANG¹ 

¹School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China

²Department of Cyberspace Security, University of Science and Technology of China, Hefei 230027, China

³The 30th Institute, China Electronics Technology Group Corporation, Chengdu 610041, China

Corresponding author: Hanyuan Zhang (zhanghanyuan1@yeah.net)

This work was supported in part by the National Natural Science Foundation of China under Grant 62003191, in part by the Natural Science Foundation of Shandong Province under Grant ZR2020QF072, and in part by the Doctoral Research Fund Project of Shandong Jianzhu University under Grant XNBS1821.


ABSTRACT Although the unsupervised extreme learning machine (UELME) based methods have been widely used to diagnosis the nonlinear process faults recently, the UELME algorithm is only designed to preserve the local adjacency similarity of the input dataset instead of mining the intra-class variations. Besides, the determination of the optimal UELME hidden nodes number is a tough issue. In order to deal with these two problems, a novel enhanced UELME (EUELME) based scheme is developed to effectively detect the nonlinear process faults in our work. In the proposed EUELME approach, the UELME algorithm is first improved by naturally incorporating the diversity analysis technique into the original UELME objective function to preserve both the intra-class variation and the local adjacency similarity of the input dataset. Then, to settle the difficult issue of selecting the optimal number of hidden nodes, kernel trick is further employed in the EUELME approach to mine the data strong nonlinearity. Based on the extracted diversity and local similarity low dimensional feature information, the k-nearest neighbor (KNN) principle is applied to derive a monitoring statistic for fault detection. At last, the experiments and comparisons on the monitoring effectiveness of the suggested EUELME based approach are made on a numerical nonlinear system and the benchmark Tennessee Eastman (TE) process. The obtained monitoring results illustrate that the significant improvements can be achieved by the proposed EUELME based fault detection approach compared with other popular and related approaches.

INDEX TERMS Nonlinear fault detection, diversity analysis, extreme learning machine, kernel trick, k-nearest neighbor principle.

I. INTRODUCTION

With the increasing demand of industrial process security and reliability, fault detection technology has been paid more and more attention. Recently, as massive measurements are stored in industrial production processes by using advanced computer control systems, the data-driven based monitoring approaches are becoming a fascinating topic and gains increasing interests. Some classical data-driven multivariate statistical approaches, such as partial least squares (PLS) and principal component analysis (PCA) based methods are broadly utilized to discover the faults [1], [2]. To deal with the

nonlinearity of the in real process data, the improved versions of the traditional PCA and PLS are discussed for nonlinear process fault detection, for instance the neural network or kernel trick based PCA [3], [4] and the kernel trick based PLS [5]. To enhance the quality of fault detection by suppressing the effect of measurement errors, Sheriff *et al.* [6] suggested a novel monitoring scheme by combining multi-scale representation based PCA with moving window generalized likelihood ratio test technology. Fezai *et al.* [7] discussed an online reduced kernel PCA based fault detection method to tackle the conventional kernel PCA's limitation of monitoring dynamic systems with large training dataset. In order to cope with the process parameters changes, measurements' errors and uncertainties over the long operation

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

periods, two improved interval reduced kernel PLS models were proposed to monitor large scale nonlinear uncertain systems in the literature [8]. Nevertheless, during the dimensionality reduction procedure, these nonlinear extensions omit the detailed local adjacency similarity structure among neighboring samples, because they only focus on the diversity information (i.e., the intra-class variations) of the samples.

As an efficient learning technique, extreme learning machine (ELM) is receiving a lot of interests in data-driven based fault detection and diagnosis fields. The ELM is indeed a single layer feed forward neural networks, where the network hidden layer parameters are experientially determined [9]. By transforming the original input dataset into a high dimensional space via the nonlinear transformation function, the ELM is an effective approach to deal with the nonlinearity of process data [10]. In order to diagnose the nonlinear process faults effectively, Boldt *et al.* [11] integrated the concept of cascade feature selection into the ELM model to combine different feature selection methods. Luo *et al.* [12] employed the real-valued gravitational search algorithm to optimize the input parameters of the ELM model, for the purpose of identifying the fault patterns of rolling element bearings. To classify the nonlinear mixed data containing numerical and categorical values, Li *et al.* [13] developed an improved radial basis function based ELM method by fusing the data process into the ELM classification. However, in the ELM's nonlinear conversion procedure, the number of optimal hidden nodes is empirically selected [14]. The number of optimal hidden nodes is a critical parameter in the ELM model because it affects the ELM's performance greatly. Therefore, the problem of using the clumsy approaches to choose the optimal number of hidden nodes should be avoided [15], [16].

The goal of the ELM is to perform the supervised mission, which immensely restrains its applicability. Nevertheless, in actual application, label data is time consuming and expensive to obtain for fully supervised learning, while the massive unlabeled data are easy to achieve. To enable the ELM to have the ability of utilizing unlabeled data, some semi-supervised and unsupervised versions are discussed. With the help of a modified loss function, Luo *et al.* [17] suggested a new semi-supervised ELM (SELM) to suppress the bad influences of outliers in labeled and unlabeled datasets. Huang *et al.* [18] discussed a distributed SELM to handle the shortcomings in the time-varying communication network. By combining with the random vector functional link networks, Peng *et al.* [19] developed a joint optimization framework based extension of SELM to use both labeled and unlabeled samples. To improve the effectiveness in disposing non-Gaussian noises, Yang *et al.* [20] proposed a novel SELM method based on robust regularized correntropy criterion. By means of integrating Laplacian regularization to learn the manifold structure of hole image samples, Lei *et al.* [21] further discussed a modified version of the SELM to classify the superheat degree.

To exploit the low dimensional features of the unlabeled data, many unsupervised ELM (UELM) extensions are proposed. In order to consider the local connectivity during graph learning, Zeng *et al.* [22] developed an adaptive locality-constrained clustering based the UELM for unsupervised learning and clustering. Huang *et al.* [23] came up with a modified UELM model for the clustering task by combining the manifold framework with the UELM algorithm. To explore the data structure much better during clustering, Peng *et al.* [24] suggested a discriminative UELM based scheme to make use of the adjacency intrinsic structure and global discriminative information of the measurements. To perform the cluster task in process data, Chen *et al.* [25] integrated UELM with L2,1 norm regularization to remove the useless hidden nodes. For the purpose of detecting abnormality in video object trajectories, Sekh *et al.* [26] combined dynamic ELM with hierarchical temporal memory together in an unsupervised way. To tackle the issue of big data, Yara and Mariette [27] suggested three improved UELM algorithms utilizing a distributed framework to learn clustering models from big data. The research works have found that the standard UELM algorithm are related to the Laplacian eigenmaps algorithm because they both first construct an affinity matrix and then utilize the spectral technique to accomplish embedding or clustering task.

The basic idea of the conventional UELM is to guarantee: in the input space, the closer the two data points are; in the output space, the more similar the two data points' predictions are. According to the principle, the conventional UELM model can be viewed as a local adjacency similarity structure analysis technique because the UELM only explores the inner relationships among different data points in the input space. Therefore, the conventional UELM algorithm only pays close attention to the detailed adjacency intrinsic structure instead of preserving the intra-class diversity or variations in a dataset. However, the omitted intra-class diversity in the UELM model is also critical for fault detection. In addition, to guarantee the efficient fault detection performance, the optimal number of UELM hidden nodes needs to be chosen, which is a troublesome and intractable task using the existing parameter selection approaches.

On the basis of the aforementioned analysis, a new monitoring approach using an enhanced UELM (EUELM) model is developed to detect the nonlinear process fault in this paper. The proposed EUELM model is constructed by integrating the intra-class diversity analysis into standard UELM model to maintain the intra-class variations of original input data. Besides, the kernel trick is introduced into the UELM to cope with the difficult problem of setting up the number of hidden nodes. The goal of the EUELM model is to preserve both the intra-class diversity information and local adjacency similarity structure of the original input data. Based on the kernel trick, the objective function EUELM is transformed into a generalized eigenvalue decomposition problem. To monitor the extracted data features using EUELM model, the

k-nearest neighbor (KNN) principle is applied to establish a fault detection statistic using the low dimensional features' local neighborhoods. The experiments and comparisons on the Tennessee Eastman (TE) process proves the superior fault detection effect of the suggested EUELM based approach.

Our work has three main contributions, which are elaborated as follows.

(1) To extract the intra-class variations and local adjacency similarity structure of the input data, the intra-class diversity analysis is infused with the traditional UELM model, which is beneficial to improve the process monitoring effect.

(2) To figure out the challenging trouble of selecting the optimal number of UELM hidden nodes explicitly, kernel trick is employed to dispose of the nonlinear property of the input data.

(3) The KNN rule is employed to establish a fault detection statistic based on the derived low dimensional features.

The remaining parts of this paper are as follows. In Section 2, the original ELM and standard UELM are reviewed briefly. The proposed EUELM model is presented in Section 3 in detail. Section 4 presents the construction of the monitoring statistic using the KNN rule. Section 5 gives an EUELM model based nonlinear process monitoring strategy. The experiments and comparisons on a numerical nonlinear system and the benchmark TE process are carried out in Section 6. At last, the conclusion is made in Section 7.

II. THE ELM AND UELM ALGORITHMS

The ELM and UELM are closely related to the proposed EUELM algorithm, so they are first reviewed to facilitate introducing EUELM model.

A. THE ELM ALGORITHM

The basic idea of original ELM [23], [28] is to calculate the output weights by using random feature mapping in matrix operations. Given N training samples $\{X, Y\} = \{x_i, y_i\}_{i=1}^N$, where $x_i \in R^{n \times 1}$ is a data point and $y_i \in R^{n_o \times 1}$ is a binary vector, n and n_o are respectively the dimensions of input layer and output layer. For y_i , only one entry corresponding to the category of x_i equals to one. The formulation of the ELM is given as

$$\sum_{i=1}^L \beta_i G(w_i \cdot x_j + b_i) = o_j, \quad j = 1, 2, \dots, N \quad (1)$$

where L and $G(w_i \cdot x + b_i)$ represent the number of hidden nodes and the activation function, respectively. $w_i = [w_{i1}, w_{i1}, \dots, w_{in}]$ indicates the input weight while $\beta_i \in R^{n_o \times 1}$ indicates the output weight. b_i denotes the bias of the i -th hidden node, while $o_j \in R^{n_o \times 1}$ is the output vector.

Given the number of hidden nodes L , Eq. (1) can be reformulated as

$$H\beta = O \quad (2)$$

where β and O are defined as $\beta = [\beta_1, \beta_2, \dots, \beta_L]^T$ and $O = [o_1, o_2, \dots, o_N]^T$. H represents the feature

transformation matrix.

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} G(w_1 \cdot x_1 + b_1) & \cdots & G(w_L \cdot x_1 + b_L) \\ \vdots & \cdots & \vdots \\ G(w_1 \cdot x_N + b_1) & \cdots & G(w_L \cdot x_N + b_L) \end{bmatrix}_{N \times L} \quad (3)$$

Given the target matrix $Y = [y_1, y_2, \dots, y_N]^T$, ELM model is designed to minimize the training error $\|O - Y\|^2$ [9]. In this regard, the ELM model learns the training samples without the residuals. Thus, there exists β such that

$$H\beta = Y \quad (4)$$

At the beginning of the learning task in the original ELM, the hidden layer bias b_i and the input weights w_i are determined according to an uniform probability distribution on $[-1, 1]$ randomly, for $i = 1, 2, \dots, L$. After the parameters w_i and b_i are selected, the β is computed as

$$\beta = H^\dagger Y \quad (5)$$

where the H^\dagger is an inverse of the H , which is established by figuring out the single value decomposition or least-squares.

In order to make the network possess good generalization property, the ELM trained network [23] is designed to obtain both the smallest training error and output weights norm [23], [29].

$$\text{Minimize } \|H\beta - Y\|^2 \text{ and } \|\beta\|^2 \quad (6)$$

On the basis of Eq. (6), the expression of ELM model is given as

$$\min_{\beta} \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \sum_{i=1}^N \|e_i\|^2$$

$$\text{s.t. } h(x_i)\beta = y_i^T - e_i^T, \quad i = 1, 2, \dots, N \quad (7)$$

where e_i denotes the i -th training sample's error vector and C indicates the penalty factor.

B. THE UELM ALGORITHM

The main objective of UELM is to guarantee that the probabilities $P(y|x_i)$ and $P(y|x_j)$ of adjacency input samples x_i and x_j should be also similar in the output space [18]. To enforce this goal, the following optimization objective is adopted.

$$\min \frac{1}{2} \sum_{i,j} w_{ij} \|P(y|x_i) - P(y|x_j)\|^2 \quad (8)$$

where w_{ij} is designed to put on a large punishment if big variation exists in the values of $P(y|x_i)$ and $P(y|x_j)$.

The weight parameter w_{ij} has the ability of representing the neighborhood relations of different samples in the original input space. The value of w_{ij} is set to be nonzero if the x_i or x_j is in the k nearest adjacencies of the x_i or x_j , respectively.

The nonzero value can be calculated by virtue of utilizing kernel function $\exp(-\|x_i - x_j\|^2 / t)$, or set to be 1. Therefore, a sparse symmetric matrix $\mathbf{W} = [w_{ij}] \in R^{N \times N}$ is established to represent the local adjacency similarity structure of the original input dataset.

In order to avert calculating the conditional probability, Eq. (8) is further approximated as

$$\min \frac{1}{2} \sum_{i,j} w_{ij} \|\hat{y}_i - \hat{y}_j\|^2 = \min \text{Tr}(\hat{\mathbf{Y}}^T \mathbf{L} \hat{\mathbf{Y}}) \quad (9)$$

where \hat{y}_i and \hat{y}_j respectively are the predictions of the input samples x_i and x_j . The symbol $\text{Tr}(\cdot)$ indicates matrix trace computation and $\hat{\mathbf{Y}}$ denotes the prediction matrix. $\mathbf{L} = \mathbf{D} - \mathbf{W}$ represents the Laplacian matrix and the element of diagonal matrix \mathbf{D} is computed as $d_{ii} = \sum_{j=1}^N w_{ij}$.

By incorporating the manifold regularization to utilize unlabeled data, the optimization of the conventional UELM model is built as

$$\begin{aligned} \min_{\beta} \frac{1}{2} \|\beta\|^2 + \frac{\lambda}{2} \text{Tr}(\mathbf{F}^T \mathbf{L} \mathbf{F}) \\ \text{s.t. } \mathbf{f}_i = \mathbf{h}(x_i)\beta, \quad i = 1, 2, \dots, N \end{aligned} \quad (10)$$

where λ indicates the tradeoff parameter and Laplacian matrix \mathbf{L} is estimated from the unlabeled training dataset. \mathbf{F} represents the output matrix whose i -th row equals to \mathbf{f}_i .

III. THE ENHANCED UELM MODEL

On the basis of Eq. (10), the optimization of standard UELM is indeed to guarantee that the corresponding network outputs f_i and f_j of the two neighboring input samples x_i and x_j are also near neighbors. Nevertheless, standard UELM model has no explicit constraint condition for distant samples in original input space. This would lead to take no account of significant variance information of original process data in the UELM model, and the faraway input samples are inclined to be projected to a small adjacent area in the output space. At this point, the UELM model is thought to be a local structure preserving algorithm, while it has no capability of mining the important global structure information of original input data. In some cases, it will project all the original input samples to one point in the output space, which makes the derived output weights overfit to the training samples. In addition, it is an intractable and troublesome task to set up the optimal number of UELM hidden nodes.

Motivated by the above analysis, an enhanced UELM (EUELM) algorithm is introduced to improve the nonlinear process fault detection effectiveness by modifying the UELM optimization with the intra-class diversity analysis technique. The primary target of EUELM is to ensure that the neighboring input samples should be projected to a small adjacent area in the latent space, while the faraway input samples should be projected to be still distinct from each other. Moreover, in order to avoid explicitly setting up the optimal hidden nodes number, kernel trick is employed to the

EUELM model, for the purpose of handling the nonlinearity of the original input data.

A. MODIFY THE UELM BY INTEGRATING THE INTRA-CLASS DIVERSITY INFORMATION

It is widely recognized that the intra-class diversity information (i.e., the intra-class variations) of the original process data [30]–[32] also contributes to achieve more efficient fault detection performance. To preserve the diversity information of input data points, a diversity graph is first defined. Then, its affinity matrix is constructed by considering the intra-class variations of data points in the diversity graph. Finally, the diversity information is efficiently maintained by maximizing the diversity scatter calculated from the diversity graph.

Given the normalized training data $\mathbf{X} = [x_1, x_2, \dots, x_N]$, the diversity graph [30]–[32] is constructed as $\mathbf{G}_d = (\mathbf{X}, \mathbf{E}, \mathbf{Q})$, where \mathbf{E} represents the set of edges connecting different data points and \mathbf{Q} denotes the affinity matrix with the elements characterizing the diversity of two different data points x_i and x_j . From the view of statistic, an element q_{ij} in the affinity matrix \mathbf{Q} can be defined as follows

$$q_{ij} = \begin{cases} 1 - \exp\left(\frac{-\|x_i - x_j\|^2}{2r^2}\right), & \text{if } x_i \text{ is not among } k \text{ nearest} \\ & \text{neighbors of } x_j \\ & \text{or } x_j \text{ is not among } k \text{ nearest} \\ & \text{neighbors of } x_i \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where q_{ij} measures the contribution of data point x_i associated with data point x_j to the diversity information.

To mine the data's diversity in the network outputs of the UELM, the objective function of preserving data diversity information is established as

$$J_D = \max \frac{1}{2} \sum_{i,j=1}^N q_{ij} \|f_i - f_j\|^2 \quad (12)$$

where $f_i = \mathbf{h}(x_i)\beta$ and $f_j = \mathbf{h}(x_j)\beta$ respectively indicate the UELM network outputs corresponding to the data points x_i and x_j .

If two data points x_i and x_j are far apart in the input space, while the corresponding network output points f_i and f_j are close to each other, then Eq. (12) will incur a heavy penalty by utilizing the defined weight q_{ij} . Therefore, maximizing Eq. (12) is intended to guarantee that if the diversity of the original input samples x_i and x_j is large, the diversity of the corresponding network output points f_i and f_j will be also large. From the perspective of statistic, the optimization in Eq. (12) enables UELM model to preserve the most input data diversity information during computing the network outputs.

The network outputs $f_i = \mathbf{h}(x_i)\beta$ and $f_j = \mathbf{h}(x_j)\beta$ are substituted into the optimization of preserving diversity defined in Eq. (12), and the objective function is reformulated

as follows

$$\begin{aligned}
 J_D(\boldsymbol{\beta}) &= \max \frac{1}{2} \sum_{i,j=1}^N q_{ij} \|\mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} - \mathbf{h}(\mathbf{x}_j)\boldsymbol{\beta}\|^2 \\
 &= \max \frac{1}{2} \sum_{i,j=1}^N q_{ij} \boldsymbol{\beta}^T (\mathbf{h}(\mathbf{x}_i) - \mathbf{h}(\mathbf{x}_j))^T (\mathbf{h}(\mathbf{x}_i) - \mathbf{h}(\mathbf{x}_j)) \boldsymbol{\beta} \\
 &= \max \left(\sum_{i,j=1}^N \boldsymbol{\beta}^T \mathbf{h}^T(\mathbf{x}_i) q_{ij} \mathbf{h}(\mathbf{x}_i) \boldsymbol{\beta} \right. \\
 &\quad \left. - \sum_{i,j=1}^N \boldsymbol{\beta}^T \mathbf{h}^T(\mathbf{x}_i) q_{ij} \mathbf{h}(\mathbf{x}_j) \boldsymbol{\beta} \right) \\
 &= \max \text{Tr} \left(\boldsymbol{\beta}^T \mathbf{H}^T \mathbf{S} \mathbf{H} \boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{H}^T \mathbf{Q} \mathbf{H} \boldsymbol{\beta} \right) \\
 &= \max \text{Tr} \left(\boldsymbol{\beta}^T \mathbf{H}^T (\mathbf{S} - \mathbf{Q}) \mathbf{H} \boldsymbol{\beta} \right) \\
 &= \max \text{Tr} \left(\boldsymbol{\beta}^T \mathbf{H}^T \mathbf{M} \mathbf{H} \boldsymbol{\beta} \right) \tag{13}
 \end{aligned}$$

where $\text{Tr}(\cdot)$ indicates the matrix trace, \mathbf{S} denotes a diagonal matrix and S_{ii} is calculated as $S_{ii} = \sum_{j=1}^N q_{ij}$. $\mathbf{M} = \mathbf{S} - \mathbf{Q}$ represents the Laplacian matrix and the weighted diversity scatter matrix is computed as $\mathbf{H}^T \mathbf{M} \mathbf{H}$.

The network output $\mathbf{f}_i = \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta}$ is also substituted into the UELM model given in Eq. (10), the UELM objective function is reformulated as

$$\begin{aligned}
 J_L(\boldsymbol{\beta}) &= \min \|\boldsymbol{\beta}\|^2 + \lambda \text{Tr}(\boldsymbol{\beta}^T \mathbf{H}^T \mathbf{L} \mathbf{H} \boldsymbol{\beta}) \\
 &= \min \text{Tr} \left(\boldsymbol{\beta}^T (\mathbf{I}_L + \lambda \mathbf{H}^T \mathbf{L} \mathbf{H}) \boldsymbol{\beta} \right) \tag{14}
 \end{aligned}$$

where $\mathbf{I}_L \in R^{L \times L}$ denotes an identity matrix.

As previously mentioned, the UELM model only pays close attention to the local adjacency similarity of the input samples, because the UELM merely keep up the data points' local neighborhood relationships. However, it is known to all that the intra-class diversity depicts the external shape of the input dataset, while the local adjacency similarity retains the input dataset's internal organization [30], [32]–[34]. Therefore, to further increase the nonlinear process monitoring effect, it is very necessary to maintain the maximal data diversity information of faraway samples as well as to keep up the adjacency similarity structure of neighboring samples.

Motivated by this, the UELM model is modified by integrating data diversity information into its standard objective function. More specifically, the EUELM optimization is constructed by virtue of maximizing the data diversity objective function $J_D(\boldsymbol{\beta})$ as well as minimizing the standard UELM objective function $J_L(\boldsymbol{\beta})$.

$$\begin{aligned}
 J_{EUELM}(\boldsymbol{\beta}) &= \frac{\min J_L(\boldsymbol{\beta})}{\max J_D(\boldsymbol{\beta})} = \frac{\min \text{Tr} \left(\boldsymbol{\beta}^T (\mathbf{I}_L + \lambda \mathbf{H}^T \mathbf{L} \mathbf{H}) \boldsymbol{\beta} \right)}{\max \text{Tr} \left(\boldsymbol{\beta}^T \mathbf{H}^T \mathbf{M} \mathbf{H} \boldsymbol{\beta} \right)} \\
 &= \min \frac{\boldsymbol{\beta}^T (\mathbf{I}_L + \lambda \mathbf{H}^T \mathbf{L} \mathbf{H}) \boldsymbol{\beta}}{\boldsymbol{\beta}^T \mathbf{H}^T \mathbf{M} \mathbf{H} \boldsymbol{\beta}} \tag{15}
 \end{aligned}$$

To reduce the computation complexity and increase the model stability, we only consider the case of $N < L$ when

computing $\boldsymbol{\beta}$ in our work [28]. The case would lead $\boldsymbol{\beta}$ to be infinite solutions. For the purpose of handling this troublesome issue, $\boldsymbol{\beta}$ is restrained to be figured out as

$$\boldsymbol{\beta} = \mathbf{H}^T \mathbf{A} \tag{16}$$

where $\mathbf{A} \in R^{N \times n_o}$ represents the loading matrix.

Then, Eq. (15) is further formulated as

$$\begin{aligned}
 J_{EUELM} &= \min \frac{\mathbf{A}^T \mathbf{H} (\mathbf{I}_L + \lambda \mathbf{H}^T \mathbf{L} \mathbf{H}) \mathbf{H}^T \mathbf{A}}{\mathbf{A}^T \mathbf{H} \mathbf{H}^T \mathbf{M} \mathbf{H} \mathbf{H}^T \mathbf{A}} \\
 &= \min \frac{\mathbf{A}^T (\mathbf{H} \mathbf{H}^T + \lambda \mathbf{H} \mathbf{H}^T \mathbf{L} \mathbf{H} \mathbf{H}^T) \mathbf{A}}{\mathbf{A}^T \mathbf{H} \mathbf{H}^T \mathbf{M} \mathbf{H} \mathbf{H}^T \mathbf{A}} \tag{17}
 \end{aligned}$$

By figuring out the following Eq. (18), the loading matrix \mathbf{A} is composed of the eigenvectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_{n_o}$ corresponding to the first n_o smallest eigenvalues $\gamma_1, \gamma_2, \dots, \gamma_{n_o}$.

$$(\mathbf{H} \mathbf{H}^T + \lambda \mathbf{H} \mathbf{H}^T \mathbf{L} \mathbf{H} \mathbf{H}^T) \boldsymbol{\alpha}_j = \gamma_j \mathbf{H} \mathbf{H}^T \mathbf{M} \mathbf{H} \mathbf{H}^T \boldsymbol{\alpha}_j \tag{18}$$

Notice that \mathbf{H} is of full row rank because of $N < L$, therefore $\mathbf{H} \mathbf{H}^T$ is invertible. We can further get

$$(\mathbf{I}_N + \lambda \mathbf{L} \mathbf{H} \mathbf{H}^T) \boldsymbol{\alpha}_j = \gamma_j \mathbf{M} \mathbf{H} \mathbf{H}^T \boldsymbol{\alpha}_j \tag{19}$$

where $\mathbf{I}_N \in R^{N \times N}$ represents the identity matrix.

B. EMPLOY THE KERNEL TRICK TO THE MODIFIED UELM MODEL

To figure out the issue of explicitly setting up the optimal modified UELM's hidden nodes number, kernel trick [35], [36] is employed. By using the kernel function $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{h}(\mathbf{x}_i), \mathbf{h}(\mathbf{x}_j) \rangle$, the kernel matrix \mathbf{K} of the proposed EUELM model is defined as follows

$$\mathbf{K} = \mathbf{H} \mathbf{H}^T : \mathbf{K}_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{h}(\mathbf{x}_i), \mathbf{h}(\mathbf{x}_j) \rangle \tag{20}$$

where $i, j = 1, 2, \dots, N$.

In our work, the Gaussian kernel is chosen as the kernel function [35]–[37].

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \sigma) \tag{21}$$

where the kernel parameter σ is set up beforehand.

Then Eq. (19) can be expressed as

$$(\mathbf{I}_N + \lambda \mathbf{L} \mathbf{K}) \boldsymbol{\alpha}_j = \gamma_j \mathbf{M} \mathbf{K} \boldsymbol{\alpha}_j \tag{22}$$

After resolving Eq. (22), the $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_{n_o}$ eigenvectors corresponding to the n_o smallest eigenvalues is achieved. At last, the value of $\boldsymbol{\beta}$ is computed as

$$\boldsymbol{\beta}^* = \mathbf{H}^T \mathbf{A} = \mathbf{H}^T [\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2, \dots, \tilde{\boldsymbol{\alpha}}_{n_o}] \tag{23}$$

where $\tilde{\boldsymbol{\alpha}}_i = \boldsymbol{\alpha}_i / \|\mathbf{H} \mathbf{H}^T \boldsymbol{\alpha}_i\| = \boldsymbol{\alpha}_i / \|\mathbf{K} \boldsymbol{\alpha}_i\|$, $i = 1, 2, \dots, n_o$ are the normalized eigenvectors and $\mathbf{A} = [\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2, \dots, \tilde{\boldsymbol{\alpha}}_{n_o}]$.

The low dimensional features of the given matrix \mathbf{X} is computed as the output matrix $\mathbf{T} \in R^{N \times n_o}$ of the EUELM model.

$$\mathbf{T} = \mathbf{H} \boldsymbol{\beta}^* = \mathbf{H} \mathbf{H}^T \mathbf{A} = \mathbf{K} \mathbf{A} \tag{24}$$

Given a test sample \mathbf{x}_t , the projection vector \mathbf{t}_t is acquired by figuring out the output vector of the EUELM model.

$$\mathbf{t}_t = \mathbf{h}(\mathbf{x}_t)\boldsymbol{\beta}^* = \mathbf{h}(\mathbf{x}_t)\mathbf{H}^T\mathbf{A} = \mathbf{k}_t\mathbf{A} \quad (25)$$

where $\mathbf{k}_t = \mathbf{h}(\mathbf{x}_t)\mathbf{H}^T \in R^{1 \times N}$ indicates the kernel vector, and $\mathbf{k}_{t,i}$ is computed as $\mathbf{k}_{t,i} = k(\mathbf{x}_t, \mathbf{x}_i)$ for $i = 1, 2, \dots, N$.

To guarantee $\sum_{i=1}^N \mathbf{h}(\mathbf{x}_i) = 0$, mean centered kernel matrix $\tilde{\mathbf{K}}$ needs to be calculated before solving Eq. (22) and Eq. (24).

$$\tilde{\mathbf{K}} = \mathbf{K} - \mathbf{I}_K\mathbf{K} - \mathbf{K}\mathbf{I}_K + \mathbf{I}_K\mathbf{K}\mathbf{I}_K \quad (26)$$

where all the elements of the $N \times N$ matrix \mathbf{I}_K are equal to $1/N$.

Before calculating the vector \mathbf{t}_t based on Eq. (25), the mean centered test kernel vector $\tilde{\mathbf{k}}_t$ also needs to be computed.

$$\tilde{\mathbf{k}}_t = \mathbf{k}_t - \mathbf{I}_t\mathbf{K} - \mathbf{k}_t\mathbf{I}_K + \mathbf{I}_t\mathbf{K}\mathbf{I}_K \quad (27)$$

where $\mathbf{I}_t = 1/N[1, \dots, 1] \in R^{1 \times N}$.

IV. FAULT DETECTION STATISTIC CONSTRUCTION BASED ON THE K-NN PRINCIPLE

After the low dimensional feature information is extracted by the EUELM model, the k-nearest neighbor (KNN) principle [38], [39] is employed to build the fault detection statistic. The basic idea of estimating the monitoring statistic utilizing the KNN principle is that a normal data point's behavior is similar to the behaviors of the training data points; while the fault data point's behavior would reveal the abnormal deviation from the behaviors of the training data points [40], [41].

That is to say, a fault data point's distances to the k nearest adjacency training data points are much bigger than that of a normal data point to the k nearest adjacency training data points. Therefore, in our work, the data point's average distances to the k nearest adjacency training data points are calculated as the fault detection statistic. After the confidence limit of the constructed monitoring statistic is determined, the test data point is thought to be normal if the average distance to its k nearest adjacency training data points is smaller than the confidence limit. Otherwise, a fault is detected.

For the given output dataset $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_N]$ of the EUELM model, the k nearest adjacency data points for each vector \mathbf{t}_i are selected in the output dataset by using the Euclidean distance.

$$d_{i,j} = \|\mathbf{t}_i - \mathbf{t}_j\|, \quad j = 1, 2, \dots, N, j \neq i \quad (28)$$

where $d_{i,j}^2$ indicates the Euclidean distance between the i -th vector \mathbf{t}_i to its j -th nearest adjacency in the output dataset. The data points owning the first k smallest Euclidean distances are selected as the k nearest adjacencies of the vector \mathbf{t}_i .

Then, the average square distance D_i^2 is computed as the fault detection statistic.

$$D_i^2 = \frac{1}{k} \sum_{j=1}^k d_{ij}^2 \quad (29)$$

To judge the status of the test data point \mathbf{x}_t , the kernel density estimation (KDE) technique [42]–[44] is applied to

estimate the threshold value D_α^2 of the monitoring statistic D_i^2 according to the output dataset \mathbf{T} .

For a test data point \mathbf{x}_t , the k nearest neighbors of its projection vector \mathbf{t}_t is also found using the following equation in the output dataset.

$$d_{t,j} = \|\mathbf{t}_t - \mathbf{t}_j\|, \quad j = 1, 2, \dots, N \quad (30)$$

Similarly, the data point owning the first k smallest values of $d_{t,j}$ are regarded as the k nearest adjacencies of the vector \mathbf{t}_t .

Then, the fault detection statistic D_t^2 of the test sample \mathbf{x}_t is calculated as

$$D_t^2 = \frac{1}{k} \sum_{j=1}^k d_{t,j}^2 \quad (31)$$

Finally, the monitoring statistic D_t^2 is compared with its corresponding confidence limit D_α^2 . If $D_t^2 > D_\alpha^2$, the test sample \mathbf{x}_t is considered as a fault sample; otherwise, the test sample \mathbf{x}_t is thought to be normal.

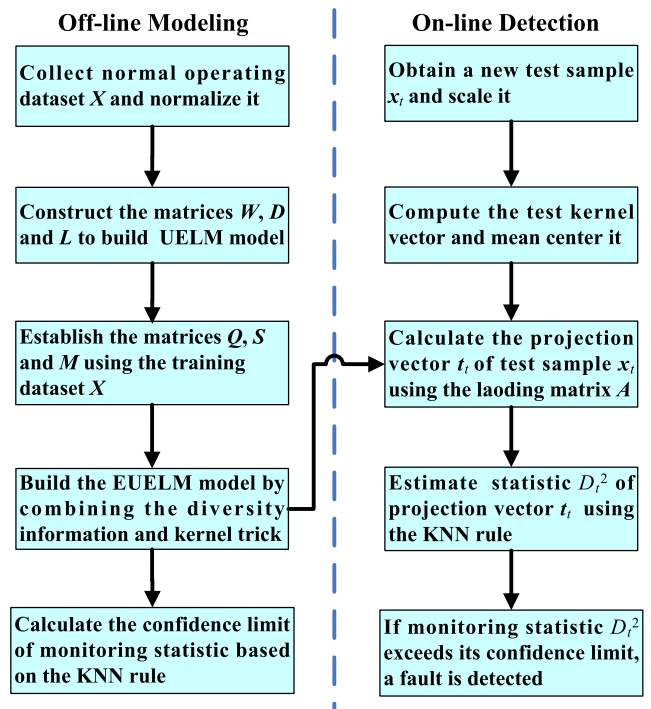


FIGURE 1. The flowchart of the proposed EUELM based fault detection approach.

V. FAULT DETECTION STRATEGY BASED ON THE EUELM MODEL

As illustrated in Fig. 1, the EUELM based monitoring approach has the off-line modelling phase and the on-line detection phase. During the former phase, the training data is used to build the EUELM model and the threshold value D_α^2 of the fault detection statistic is estimated by applying the K-NN principle to quantize the similarity between each sample and the training dataset. During the latter phase, the fault detection statistic D_t^2 of the test sample is calculated to judge

TABLE 1. The pseudo code of the EUELM based fault detection method.

Algorithm: The EUELM based fault detection for the nonlinear process
Input: The normalized normal operating dataset X and the test sample x_t .
Procedure:
<ul style="list-style-type: none"> • The off-line modelling phase <ol style="list-style-type: none"> 1. Compute the weight matrix W using the training matrix X, then figure out the matrices D ($D_{ij} = \sum_j W_{ij}$) and L ($L = D - W$). 2. Build the affinity matrix Q using the training matrix X, then compute the matrices S ($S_{ii} = \sum_{j=1}^N q_{ij}$) and M ($M = S - Q$). 3. Compute the kernel matrix K based on Eq. (20), then carry out the mean centring operation to ensure $\sum_{i=1}^N h(x_i) = 0$ using Eq. (26) and the centered kernel matrix \tilde{K} is obtained. 4. Figure out the eigenvalue problem defined in Eq. (22) to acquire the loading matrix $A = [\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_{n_o}]$, where $\tilde{a}_i = \alpha_i / \ \tilde{K} \alpha_i\$, $i = 1, 2, \dots, n_o$. 5. Compute the output matrix T according to Eq. (24), then perform the KNN rule on the matrix T to construct the monitoring statistic D^2. 6. Employ the KDE technique to the values of the monitoring statistic D^2 in order to estimate the threshold value D_α^2. • The on-line detection phase <ol style="list-style-type: none"> 1. Normalize the test sample x_t according to the training dataset X. 2. Figure out the test kernel vector k_t and mean center it according to Eq. (27) to get the mean centered test kernel vector \tilde{k}_t. 3. Establish the test projection vector t_t based on Eq. (25), then obtain the monitoring statistic D_t^2 according to Eq. (31). 4. Contrast the value of the D_t^2 with the threshold value D_α^2. If the D_t^2 exceeds the D_α^2, a fault is alarmed; otherwise, the operating status of the process is normal.
Output: The operating status of the monitoring nonlinear process.

whether the fault occurs. The detailed pseudo code of the EUELM based monitoring scheme is summarized in Table 1.

VI. CASE STUDIES

In our work, two case studies are adopted to estimate the fault detection performance of the proposed EUELM based approach. One case study is a numerical nonlinear system, and the other one is the Tennessee Eastman (TE) process which is a well-known nonlinear process. The performance comparisons with other related methods are further conducted to testify the superior fault detection capability of the EUELM based approach.

A. CASE STUDY ON A NUMERICAL NONLINEAR SYSTEM

1) PROCESS DESCRIPTION

A numerical nonlinear system involving three process variables is first adopted to testify the effect of the EUELM based fault detection approach. The utilized numerical nonlinear

system formulated in Eq. (32) is an improved version of one discussed in the literatures [45] and [46].

$$\begin{cases} x_1 = t + e_1 \\ x_2 = x_1^2 - 3x_1 + e_2 \\ x_3 = x_1^2 + 3x_2 + e_3 \end{cases} \quad (32)$$

where $e_1, e_2, e_3 \in N(0, 0.01)$ represent the noises, and t indicates the random variable sampling from $[0, 2]$. In this numerical nonlinear system, the monitored variables are the output variables $[x_1, x_2, x_3]$. Based on Eq. (32), 300 normal samples are produced to establish the training dataset. A fault is introduced to generate the test dataset possessing 300 samples. The introduced fault is set up as: variable x_1 is increased from the 51-th sample by adding $0.05 \times (k - 50)$ to its previous value until the end of the simulation.

2) COMPARATIVE METHODS AND PARAMETER SETTING

In this study, the monitoring feasibility and effect of the EUELM is contrasted with the traditional UELM and the KPCA. To be fair, for the standard UELM model, the KNN principle is also employed to the output dataset to construct the fault detection statistic.

TABLE 2. The values of the used parameters in the KPCA, UELM and EUELM.

Method \ Parameter	KPCA	UELML	EUELML
The output space dimension n_o	-	50	50
The number of hidden nodes L	-	1400	-
The kernel parameter σ	250	-	250
The number of nearest neighbors k	-	6	6
The tradeoff parameter λ	-	0.3	0.3
The number of principal components	95% variance	95% variance	95% variance

For the EUELM, the kernel function is chosen as the Gaussian kernel [35]–[37]. The kernel parameter σ and the output space dimension n_o are respectively set up as 250 and 50 according to the grid search algorithm [47], [48] by seeking the optimal fault detection result of the training dataset. The nearest neighbor number k of the KNN rule is empirically set to be 6, and the tradeoff parameter λ is determined as 0.3 with the help of the grid search. The values of the used parameters in the EUELM based fault detection method are given in Table 2. For the sake of fairness, the Gaussian kernel is also employed to the KPCA, and the kernel function is determined as 250 as well. For the UELM model, the dimension of output space n_o , the nearest neighbors number k and the tradeoff parameter λ are also respectively selected as 50, 6 and 0.3. Furthermore, the number of hidden nodes L is set up as 1400 and the activation function is the Sigmoid function. For all the three methods, the principal components possessing 95% variance of the training dataset are retained,

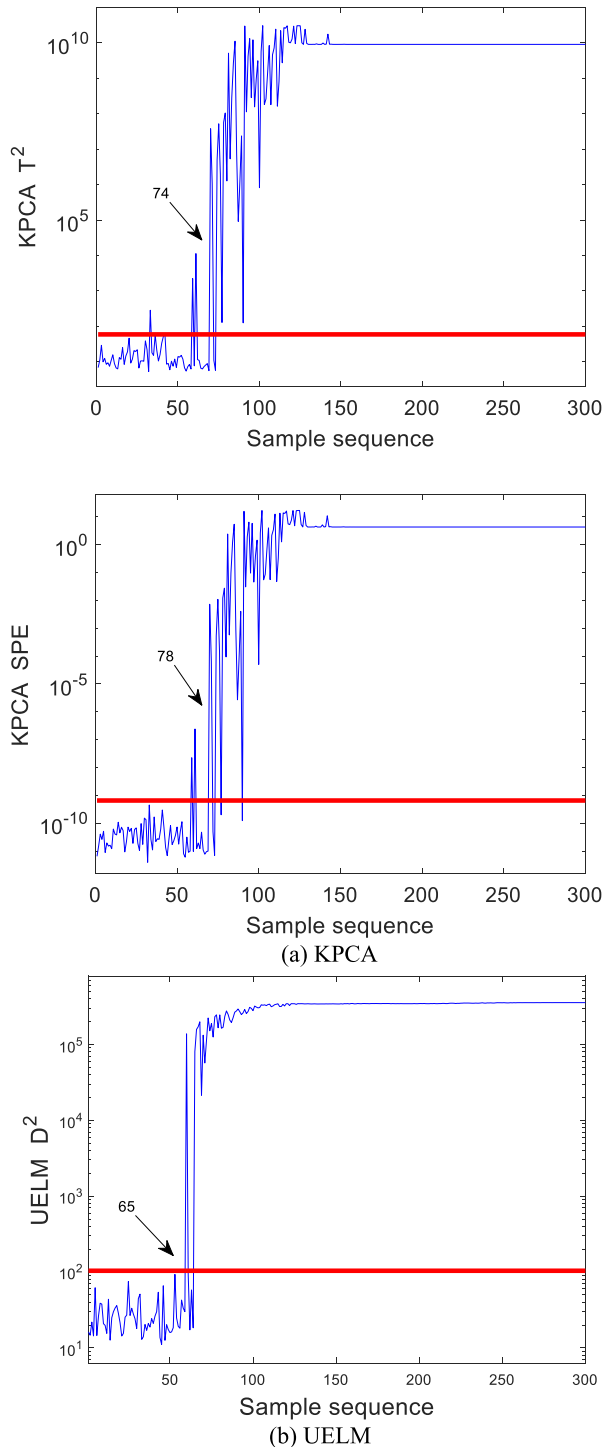


FIGURE 2. The monitoring charts of the three approaches for the simulated fault, (a) KPCA, (b) UELM and (c) EUELM.

and the threshold values are decided according to the 99% confidence level. A fault is alarmed if consecutive 5 samples go beyond the corresponding threshold value and the fault detection time is determined as the first sample number of them. The fault detection rate is computed as the percentage of the detected fault samples in the overall real fault samples.

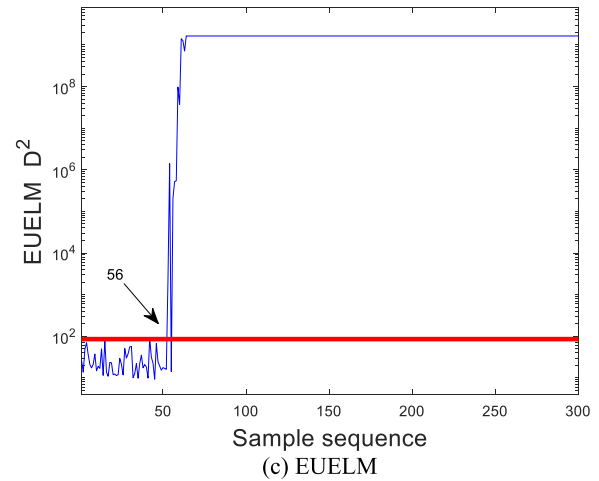


FIGURE 2. (Continued.) The monitoring charts of the three approaches for the simulated fault, (a) KPCA, (b) UELM and (c) EUELM.

TABLE 3. The fault detection times (FDTs) and fault detection rates (FDRs) of the KPCA, UELM and EUELM for the simulated fault.

Method	Statistic	FDT (Sample No.)	FDR (%)
KPCA	T^2	74	90.65
	SPE	78	87.00
UEL	D^2	65	94.31
EUELM	D^2	56	98.00

3) FAULT DETECTION EFFECT COMPARISON

The fault detection charts of KPCA, UELM and EUELM for the ramp fault are illustrated in Fig. 2. Fig. 2(a) shows that the KPCA T^2 statistic detects the process fault at the 74-th sample and its SPE statistic discovers the process fault at the 78-th sample. Compared with the KPCA, the UELM gains a better monitoring result given in Fig. 2(b), where its D^2 statistics gives an alarm of the fault at the 65-th sample with the 94.31% fault detection rate. As illustrated in Fig. 2(c), the EUELM D^2 statistic warns of the fault at the 56-th sample. Therefore, the EUELM is the most sensitive approach to the ramp fault among these three fault detection approaches. Table 3 lists the fault detection times and fault detection rates of the KPCA, the UELM and the EUELM for the simulated ramp fault. According to Table 3, the EUELM D^2 statistic has the highest fault detection rate, i.e., 98.00%. Besides, the fault detection times and fault detection rates of the EUELM, UELM and KPCA for the ramp fault are respectively visualized in Fig. 3 and Fig. 4 for a more intuitive comparison. To summarize, the experiments and comparisons on the numerical nonlinear system demonstrates the superior fault detection performance of the EUELM based approach over the KPCA and UELM based approaches.

B. CASE STUDY ON THE TE PROCESS

1) PROCESS DESCRIPTION

The TE process is a benchmark to compare various fault detection approaches [49]–[51]. The TE process is set up

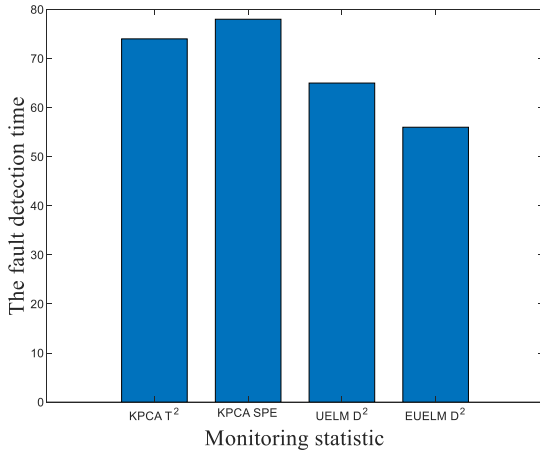


FIGURE 3. The fault detection times of the KPCA, UELM and EUELM for the simulated fault.

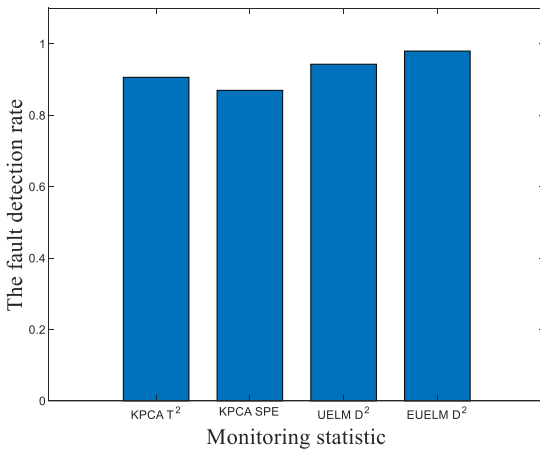


FIGURE 4. The fault detection rates of the KPCA, UELM and EUELM for the simulated fault.

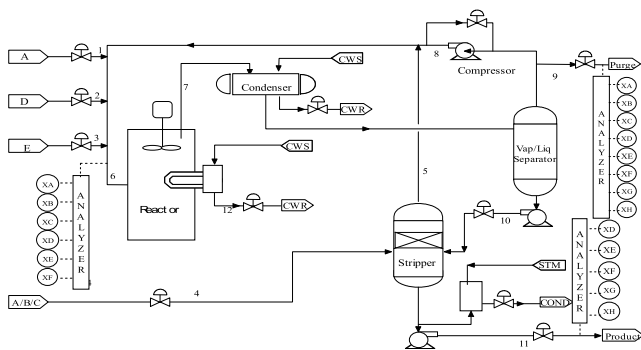


FIGURE 5. The TE process flowchart.

on the basis of a plant-wide industrial chemical operation model given in Fig. 5, which is composed of reactor, stripper, separator, condenser, and compressor. According to the references [51] and [52], 52 important variables are selected as the monitoring variables. A TE simulator can be found in the website: <http://brahms.scs.uiuc.edu>, which permits one normal operating mode and 21 fault patterns. Note that the fault patterns IDV(3), IDV(9), IDV(15) and IDV(19) have been

TABLE 4. The fault patterns of the TE process [51], [52].

Fault ID	Fault variable	Fault type
IDV(1)	A/C feed ratio, B composition constant (stream 4)	Step
IDV(2)	B composition, A/C feed ratio constant (stream 4)	Step
IDV(4)	Reactor cooling water inlet temperature	Step
IDV(5)	Condenser cooling water inlet temperature	Step
IDV(6)	A feed loss (stream 1)	Step
IDV(7)	C header pressure loss-reduced availability (stream 4)	Step
IDV(8)	A, B and C feed compositions (stream 4)	Random variation
IDV(10)	C feed temperature (stream 4)	Random variation
IDV(11)	Reactor cooling water inlet temperature	Random variation
IDV(12)	Condenser cooling water inlet temperature	Random variation
IDV(13)	Reaction kinetics	Slow shift
IDV(14)	Reactor cooling water valve	Sticking
IDV(16)~(18)	Unknown	Unknown
IDV(20)	Unknown	Unknown
IDV(21)	Valve position constant (stream 4)	Constant position

TABLE 5. The values of the used parameters in the three monitoring methods.

Parameter	Method		
	KPCA	UELM	EUELM
The output space dimension n_o	-	40	40
The number of hidden nodes L	-	1000	-
The kernel parameter σ	600	-	600
The number of nearest neighbors k	-	4	4
The tradeoff parameter λ	-	0.1	0.1
The number of principal components	95% variance	95% variance	95% variance

already proved to be difficultly detected by the data-driven based fault detection approaches because these fault datasets have no observable changes in the means or the variances [50], [52]. Therefore, except for these four fault patterns, the rest of the seventeen fault patterns given in Table 4 are utilized to testify the monitoring capability of the EUELM based scheme in our work. The TE simulator generates 960 samples for the normal operating mode and each introduced fault pattern. At the 160-th sample, all the seventeen faults are added to the TE process. More details about the introduction of the TE process can refer to the literature [52].

2) COMPARATIVE METHODS AND PARAMETER SETTING

In our work, the fault detection performance of the EUELM is also contrasted with the UELM and KPCA. To be fair, the KNN principle is adopted to the output dataset of the standard UELM model to establish the fault detection statistic.

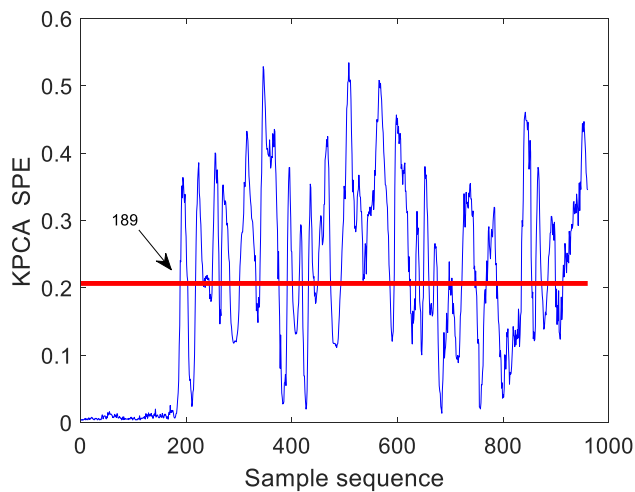
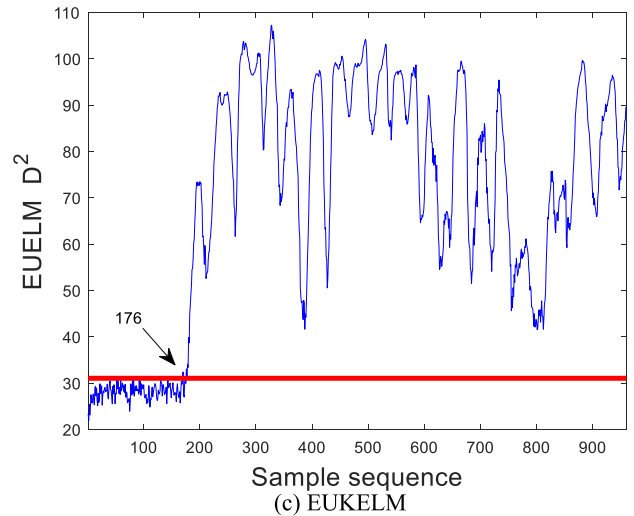
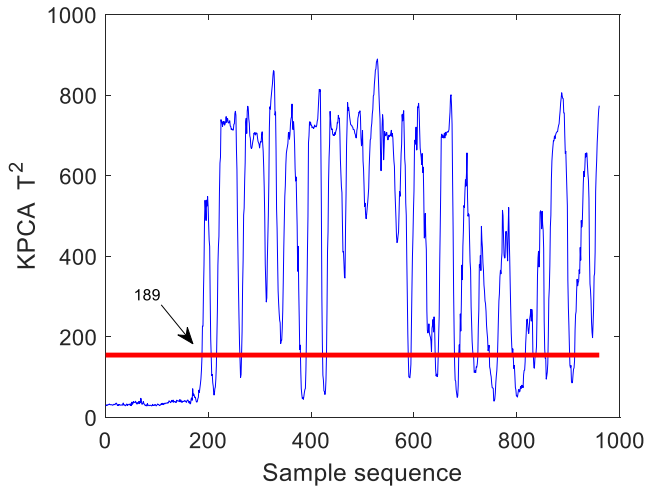
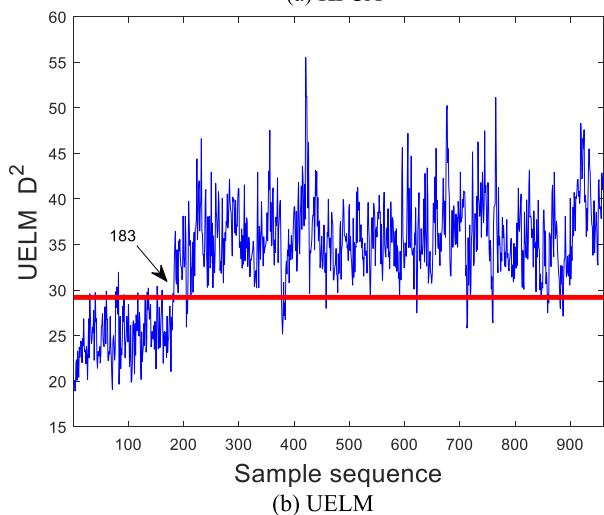


FIGURE 6. (Continued.) The monitoring charts of the three approaches for fault IDV(8), (a) KPCA, (b) UELM and (c) EUELM.

dimension n_o in the EUELM are respectively set to be 600 and 40 based on the grid search algorithm [47], [48]. The nearest neighbor number k of the KNN rule used in the EUELM is empirically chosen to be 4, and the tradeoff parameter λ is determined as 0.1 using the grid search. The values of the used parameters in the EUELM based method for the TE process are given in Table 5. To be fair, the Gaussian kernel is also utilized in the KPCA based method, and the kernel function is chosen as 600 as well. In the UELM based approach, the dimension of output space n_o , the nearest neighbors number k and the tradeoff parameter λ are also set to be 40, 4 and 0.1, respectively. Moreover, the number of hidden nodes L is set up as 1000 in the UELM model and the activation function is the Sigmoid function. For all the three monitoring methods, the principal components possessing 95% variance of the training dataset are retained, and the threshold values are decided in line with the 99% confidence level. A fault is alarmed if consecutive 5 samples go beyond the corresponding threshold value and the fault detection time is determined as the first sample number of them. The fault detection rate is computed as the percentage of the detected fault samples in the overall real fault samples.

(a) KPCA



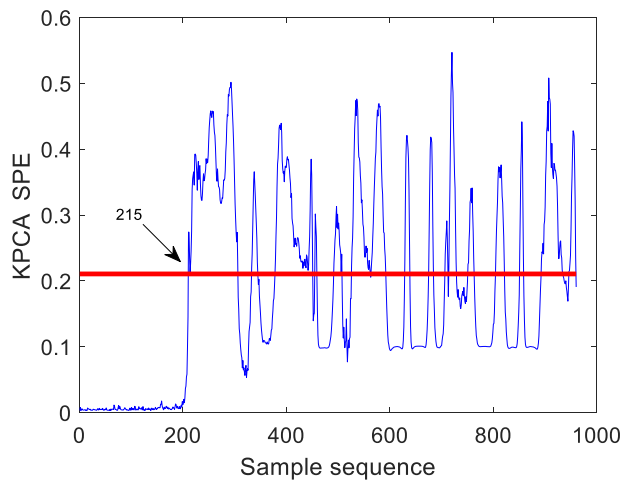
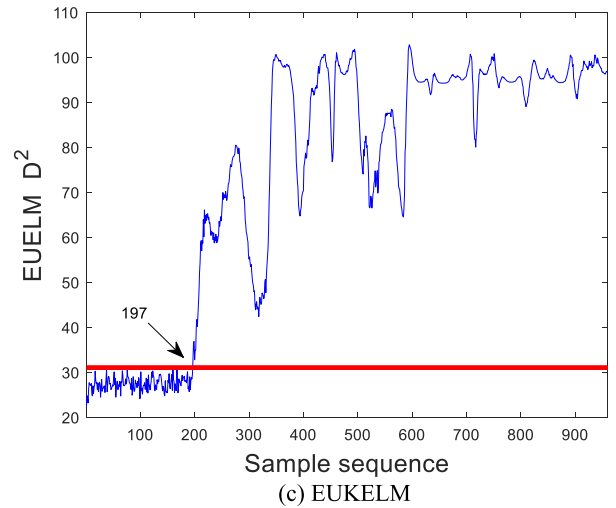
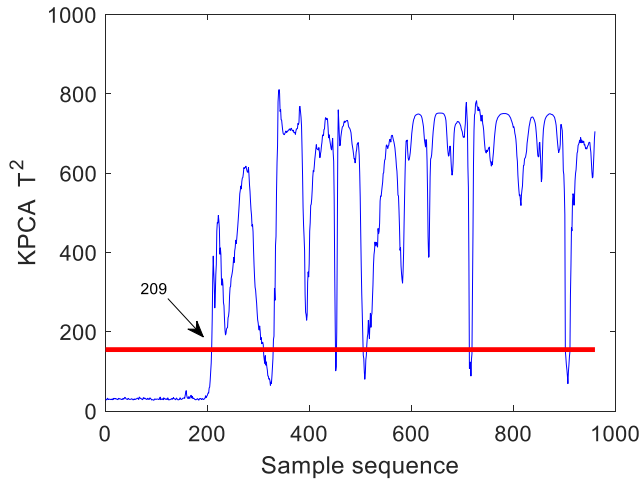
(b) UELM

FIGURE 6. The monitoring charts of the three approaches for fault IDV(8), (a) KPCA, (b) UELM and (c) EUELM.

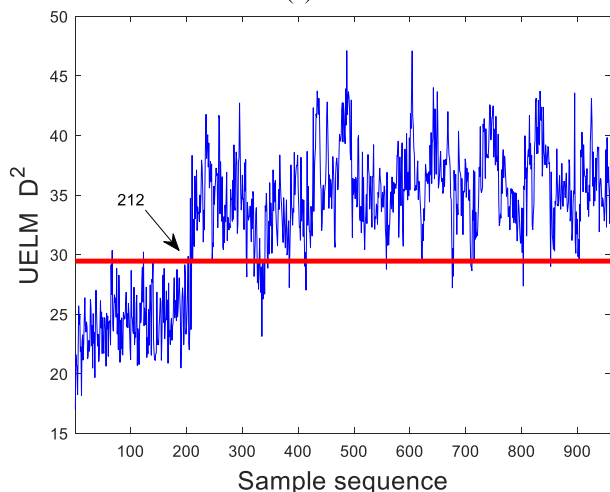
For the EUELM based scheme, the Gaussian kernel is selected as the kernel function according to the references [35], [36] and [37]. The kernel parameter σ and the output space

3) FAULT DETECTION EFFECT COMPARISON

The monitoring results for the faults IDV(8), IDV(13) and IDV(17) are utilized to confirm the superior monitoring effect of the EUELM based method. The monitoring charts of the EUELM, UELM and KPCA based approaches for the fault IDV(8) are shown in Fig. 6. From the fault detection charts of the KPCA shown in Fig. 6(a), we can see that both the T^2 and SPE statistics detect the fault at the 189-th sample. However, these two statistics both have a low fault detection rates because a lot of fault samples go down below the corresponding threshold value after the 189-th sample. From Fig. 6(b), the D^2 statistic of the UELM gives better fault detection result compared with that of the KPCA. According to Fig. 6(b),



(a) KPCA



(b) UELM

FIGURE 7. The monitoring charts of the three approaches for fault IDV(13), (a) KPCA, (b) UELM and (c) EUELM.

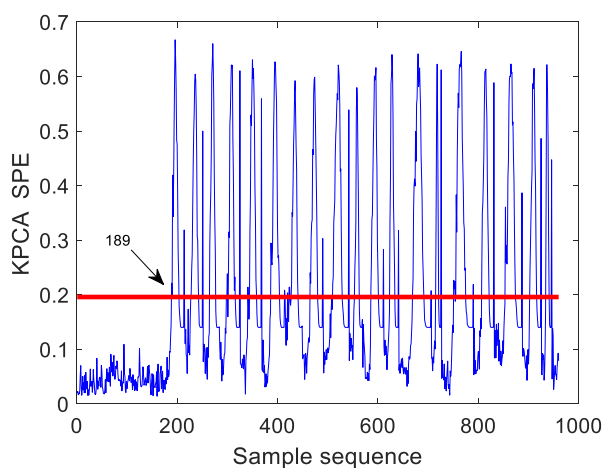
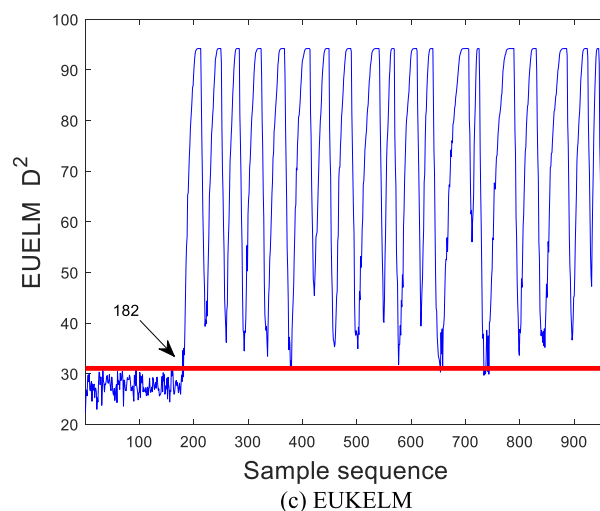
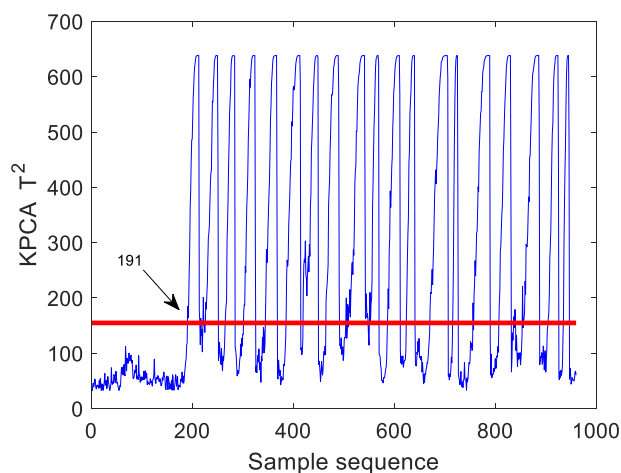
the UELM D^2 statistic discovers the fault at the 183-th sample and much fewer fault samples go down the confidence limit after the fault is detected. However, the UELM has the false alarming samples under the normal operating status.

FIGURE 7. (Continued.) The monitoring charts of the three approaches for fault IDV(13), (a) KPCA, (b) UELM and (c) EUELM.

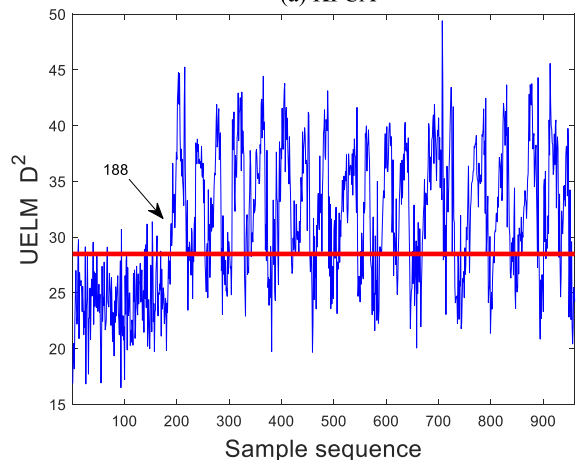
Through comparing the monitoring results of the three approaches, the EUELM achieves the best fault detection effect illustrated in Fig. 6(c). From Fig. 6(c), the D^2 statistic exceeds its threshold at the 176-th sample with no missing alerted fault samples, which reveals that the EUELM has the earliest and the most accurate fault detection results. Based on these monitoring results, the EUELM based approach is the most effective one for detecting the fault IDV(8) among the three fault detection approaches.

The monitoring charts obtained by the three monitoring approaches for the fault IDV(13) which is the slow shift in reaction kinetics are plotted in Fig. 7. In Fig. 7(a), the T^2 and SPE statistics of the KPCA respectively give an alarm of the fault IDV(13) at the 209-th sample and the 215-th sample. However, these two statistics owns low fault detection rates because many real fault samples are wrongly treated as the normal samples without giving fault alarms. As shown in Fig. 7(b), the UELM D^2 statistic alarms the fault IDV(13) at the 212-th sample. Whereas, the fault detection rate of the UELM is still needed to be improved because many fault samples also go down the threshold value after the 212-th sample. On the contrary, the EUELM results in the best fault detection performance in Fig. 7(c), where its D^2 statistic alerts the fault at the 197-th sample with the highest fault detection rate. The monitoring results of fault IDV(13) illustrate that the EUELM detects fault IDV(13) much faster and more accurate than the KPCA and UELM.

Fig. 8 illustrates the fault detection charts of the three methods for fault IDV(17). From Fig. 8(a), the KPCA has the worst monitoring performance, where its T^2 statistic discovers the fault at the 191-th sample while its SPE statistic warns the fault at the 189-th sample. Besides, both the T^2 and SPE statistics result in much lower fault detection rates because more fault samples fluctuate around the confidence limit after the fault is detected. Compared with the KPCA, the UELM brings about a slight improvement in terms of the



(a) KPCA



(b) UELM

FIGURE 8. The monitoring charts of the three approaches for fault IDV(17), (a) KPCA, (b) UELM and (c) EUELM.

monitoring performance in Fig. 8(b). According to Fig. 8(b), the UELM’s D^2 statistic exceeds the confidence limit at the 188-th sample. In addition, because of fewer missing alarmed fault samples, the D^2 statistic acquire a slight higher fault detection rate. In contrast to the results of the KPCA and UELM, the EUELM D^2 statistic given in Fig. 8(c) reacts

FIGURE 8. (Continued.) The monitoring charts of the three approaches for fault IDV(17), (a) KPCA, (b) UELM and (c) EUELM.

TABLE 6. The fault detection times (sample no.) of the KPCA, UELM and EUELM.

Fault ID	KPCA		UEL M	EUELM
	T^2	SPE	D^2	D^2
IDV(1)	167	168	165	162
IDV(2)	191	185	172	171
IDV(4)	166	161	166	161
IDV(5)	175	176	161	161
IDV(6)	167	165	161	161
IDV(7)	162	161	165	161
IDV(8)	189	189	183	176
IDV(10)	218	195	169	179
IDV(11)	167	171	171	166
IDV(12)	167	167	167	163
IDV(13)	209	215	212	197
IDV(14)	166	162	169	161
IDV(16)	356	451	348	356
IDV(17)	191	189	188	182
IDV(18)	259	264	173	238
IDV(20)	245	246	237	241
IDV(21)	676	671	482	415

the most quickly to fault IDV(17) because it goes beyond the threshold value after the 182-th sample with no missing alarmed fault samples. Besides, the EUELM achieves the highest fault detection rate, i.e., 94.47%. This again demonstrates the superior fault detection effect of the EUELM based method over the UELM and KPCA based methods.

As listed in Table 6 and Table 7, the process monitoring performance of the EUELM, UELM and KPCA based approaches for all the seventeen fault patterns are investigated. According to the Table 6, we find the UELM owns much earlier fault detection times for the faults IDV(1), IDV(2), IDV(5), IDV(6), IDV(8), IDV(10), IDV(13) and

TABLE 7. The fault detection rates (%) of the KPCA, UELM and EUELM.

Fault ID	KPCA		UEL	EUELM
	T^2	SPE	D^2	D^2
IDV(1)	99.25	86.43	99.50	99.88
IDV(2)	96.23	96.98	98.62	98.75
IDV(4)	25.50	90.07	50.43	100.00
IDV(5)	19.98	17.71	20.86	25.28
IDV(6)	99.25	99.50	100.00	100.00
IDV(7)	93.47	98.24	69.45	100.00
IDV(8)	74.50	47.86	86.81	98.12
IDV(10)	34.04	39.67	38.30	48.06
IDV(11)	51.69	25.66	22.78	65.96
IDV(12)	96.73	91.58	88.24	98.62
IDV(13)	85.93	41.58	83.54	95.48
IDV(14)	83.29	91.71	48.81	100.00
IDV(16)	14.70	17.84	23.28	23.40
IDV(17)	49.87	24.62	52.14	94.47
IDV(18)	87.69	84.67	86.71	90.36
IDV(20)	26.63	25.88	40.68	53.44
IDV(21)	31.91	32.16	35.17	44.81

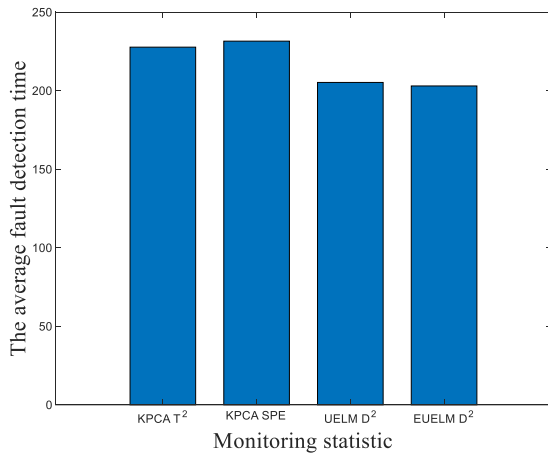


FIGURE 9. The average fault detection times of the KPCA, UELM and EUELM for the seventeen fault patterns.

IDV(16) ~ IDV(21) than that of the KPCA. For the rest of the faults apart from the faults IDV(7) and IDV(14), both the UELM and KPCA based methods have similar fault detection times. However, the monitoring capability of the UELM is still not satisfied. On the contrary, the EUELM based method gives out an improved fault detection performance. To be specific, the EUELM achieves the earliest fault detection times among these three methods for the thirteen fault patterns except for the faults IDV(10), IDV(16), IDV(18) and IDV(20). In the light of Table 7, the UELM and KPCA have the similar fault detection rates for the faults IDV(1), IDV(2) and IDV(6). Nevertheless, the UELM reveals much higher fault detection rates than that of the KPCA for the rest of the faults apart from the faults IDV(4), IDV(7), IDV(11) ~ IDV(13), IDV(14) and IDV(18). By further comparing the monitoring results of these three methods again,

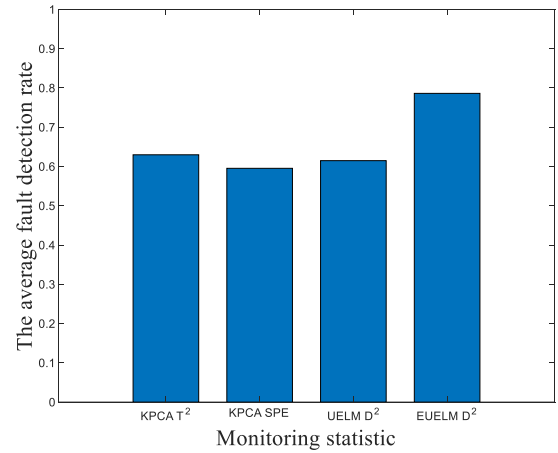


FIGURE 10. The average fault detection rates of the KPCA, UELM and EUELM for the seventeen fault patterns.

the EUELM gains the highest fault detection rates for all the seventeen fault patterns. To make a more intuitive comparison, the average fault detection times and average fault detection rates of the EUELM, UELM and KPCA over the seventeen fault patterns are visualized in Fig. 9 and Fig. 10, respectively. As illustrated in Fig. 6 and Fig. 7, the EUELM achieves the earliest average fault detection time and the highest average fault detection rate among the three methods. In brief, the comprehensive and visualized comparisons revealed in Table 6, Table 7, Fig. 9 and Fig.10 verify the excellent fault detection effect of the EUELM over the UELM and KPCA.

VII. CONCLUSION

A novel enhanced UELM based monitoring scheme is proposed to detect the nonlinear process fault in this paper. Our work has three main contributions. Firstly, to preserve both the intra-class variations and the local adjacency similarity structure of the input dataset, an improved UELM algorithm is put forward by uniting the intra-class diversity analysis technique with the conventional UELM model. Secondly, to tackle the difficult trouble of figuring out the optimal number of hidden nodes, the improved UELM model is further enhanced by applying the kernel trick to mine the data nonlinear characteristic. Thirdly, when the intra-class diversity and local adjacency similarity information of the measurements is exploited using the proposed EUELM model, the KNN rule is employed to build a fault detection statistic. Through the detailed comparisons with the traditional KPCA and closely related UELM based monitoring methods, the experimental results obtained from a numerical nonlinear system and the benchmark TE process clearly testify the superior nonlinear fault detection effect of the suggested EUELM based scheme, in terms of the fault detection time and fault detection rate.

REFERENCES

[1] H. Safaeipour, M. Forouzanfar, and A. Casavola, "A survey and classification of incipient fault diagnosis approaches," *J. Process Control*, vol. 97, pp. 1–16, Jan. 2021.

- [2] C. Shang, S. X. Ding, and H. Ye, "Distributionally robust fault detection design and assessment for dynamical systems," *Automatica*, vol. 125, Mar. 2021, Art. no. 109434.
- [3] M. Qaraei, S. Abbaasi, and K. Ghiasi-Shirazi, "Randomized non-linear PCA networks," *Inf. Sci.*, vol. 545, pp. 241–253, Feb. 2021.
- [4] W. Bounoua and A. Bakdi, "Fault detection and diagnosis of nonlinear dynamical processes through correlation dimension and fractal analysis based dynamic kernel PCA," *Chem. Eng. Sci.*, vol. 229, Jan. 2021, Art. no. 116099.
- [5] P. Zhou, R. Zhang, M. Liang, J. Fu, H. Wang, and T. Chai, "Fault identification for quality monitoring of molten iron in blast furnace ironmaking based on KPLS with improved contribution rate," *Control Eng. Pract.*, vol. 97, Apr. 2020, Art. no. 104354.
- [6] M. Z. Sherif, M. Mansouri, M. N. Karim, H. Nounou, and M. Nounou, "Fault detection using multiscale PCA-based moving window GLRT," *J. Process Control*, vol. 54, pp. 47–64, Jun. 2017.
- [7] R. Fezai, M. Mansouri, O. Taouali, M. F. Harkat, and N. Bouguila, "Online reduced kernel principal component analysis for process monitoring," *J. Process Control*, vol. 61, pp. 1–11, Jan. 2018.
- [8] R. Fezai, K. Abodayeh, M. Mansouri, A. Kouadri, M.-F. Harkat, H. Nounou, M. Nounou, and H. Messaoud, "Reliable fault detection and diagnosis of large-scale nonlinear uncertain systems using interval reduced kernel PLS," *IEEE Access*, vol. 8, pp. 78343–78353, Aug. 2020.
- [9] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 879–892, Jul. 2006.
- [10] A. Iosifidis, A. Tefas, and I. Pitas, "Minimum class variance extreme learning machine for human action recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 11, pp. 1968–1979, Nov. 2013.
- [11] F. de Assis Boldt, T. W. Rauber, and F. M. Varejão, "Cascade feature selection and ELM for automatic fault diagnosis of the tennessee eastman process," *Neurocomputing*, vol. 239, pp. 238–248, May 2017.
- [12] M. Luo, C. Li, X. Zhang, R. Li, and X. An, "Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings," *ISA Trans.*, vol. 65, pp. 556–566, Nov. 2016.
- [13] Q. Li, Q. Xiong, S. Ji, Y. Yu, C. Wu, and H. Yi, "A method for mixed data classification base on RBF-ELM network," *Neurocomputing*, vol. 431, pp. 7–22, Mar. 2021.
- [14] X. Bi, X. Zhao, G. Wang, P. Zhang, and C. Wang, "Distributed extreme learning machine with kernels based on MapReduce," *Neurocomputing*, vol. 149, pp. 456–463, Feb. 2015.
- [15] A. Iosifidis, A. Tefas, and I. Pitas, "On the kernel extreme learning machine classifier," *Pattern Recognit. Lett.*, vol. 54, pp. 11–17, Mar. 2015.
- [16] X. Liu, L. Wang, G.-B. Huang, J. Zhang, and J. Yin, "Multiple kernel extreme learning machine," *Neurocomputing*, vol. 149, pp. 253–264, Feb. 2015.
- [17] X. Luo, F. Liu, S. Yang, X. Wang, and Z. Zhou, "Joint sparse regularization based sparse semi-supervised extreme learning machine (S3ELM) for classification," *Knowl.-Based Syst.*, vol. 73, pp. 149–160, Jan. 2015.
- [18] G.-B. Huang, S. Song, J. N. D. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2405–2417, Dec. 2014.
- [19] Y. Peng, Q. Li, W. Kong, F. Qin, J. Zhang, and A. Cichocki, "A joint optimization framework to semi-supervised RVFL and ELM networks for efficient data classification," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106756.
- [20] J. Yang, J. Cao, T. Wang, A. Xue, and B. Chen, "Regularized corentropy criterion based semi-supervised ELM," *Neural Netw.*, vol. 122, pp. 117–129, Feb. 2020.
- [21] Y. Lei, X. Chen, M. Min, and Y. Xie, "A semi-supervised Laplacian extreme learning machine and feature fusion with CNN for industrial superheat identification," *Neurocomputing*, vol. 381, pp. 186–195, Mar. 2020.
- [22] Y. Zeng, J. Chen, Y. Li, Y. Qing, and G.-B. Huang, "Clustering via adaptive and locality-constrained graph learning and unsupervised ELM," *Neurocomputing*, vol. 401, pp. 224–235, Aug. 2020.
- [23] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.
- [24] Y. Peng, W.-L. Zheng, and B.-L. Lu, "An unsupervised discriminative extreme learning machine and its applications to data clustering," *Neurocomputing*, vol. 174, pp. 250–264, Jan. 2016.
- [25] J. Chen, Y. Zeng, Y. Li, and G.-B. Huang, "Unsupervised feature selection based extreme learning machine for clustering," *Neurocomputing*, vol. 386, pp. 198–207, Apr. 2020.
- [26] A. A. Sekh, D. P. Dogra, S. Kar, P. P. Roy, and D. K. Prasad, "ELM-HTM guided bio-inspired unsupervised learning for anomalous trajectory classification," *Cognit. Syst. Res.*, vol. 63, pp. 30–41, Oct. 2020.
- [27] Y. Rizk and M. Awad, "On the distributed implementation of unsupervised extreme learning machines for big data," *Procedia Comput. Sci.*, vol. 53, pp. 167–174, Jul. 2015.
- [28] G. Feng, G.-B. Huang, Q. Lin, and R. Gay, "Error minimized extreme learning machine with growth of hidden nodes and incremental learning," *IEEE Trans. Neural Netw.*, vol. 20, no. 8, pp. 1352–1357, Aug. 2009.
- [29] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 525–536, Mar. 1998.
- [30] Q. Hua, L. Bai, X. Wang, and Y. Liu, "Local similarity and diversity preserving discriminant projection for face and handwriting digits recognition," *Neurocomputing*, vol. 86, pp. 150–157, Jun. 2012.
- [31] Q.-X. Gao, D.-Y. Xie, H. Xu, Y.-Z. Li, and X.-Q. Gao, "Supervised feature extraction based on information fusion of local structure and diversity information," *Acta Automatica Sinica*, vol. 36, no. 8, pp. 1107–1114, Aug. 2010.
- [32] Q.-X. Gao, H. Xu, Y.-Y. Li, and D.-Y. Xie, "Two-dimensional supervised local similarity and diversity projection," *Pattern Recognit.*, vol. 43, no. 10, pp. 3359–3363, Oct. 2010.
- [33] H. Zhang, X. Deng, Y. Zhang, C. Hou, C. Li, and Z. Xin, "Nonlinear process monitoring based on global preserving unsupervised kernel extreme learning machine," *IEEE Access*, vol. 7, pp. 106053–106064, Aug. 2019.
- [34] M. Zhang, Z. Ge, Z. Song, and R. Fu, "Global-local structure analysis model and its application for fault detection and identification," *Ind. Eng. Chem. Res.*, vol. 50, no. 11, pp. 6837–6848, Apr. 2011.
- [35] X. Deng, P. Cai, Y. Cao, and P. Wang, "Two-step localized kernel principal component analysis based incipient fault diagnosis for nonlinear industrial processes," *Ind. Eng. Chem. Res.*, vol. 59, no. 13, pp. 5956–5968, Apr. 2020.
- [36] M. Yao and H. Wang, "On-line monitoring of batch processes using generalized additive kernel principal component analysis," *J. Process Control*, vol. 28, pp. 56–72, Apr. 2015.
- [37] L. M. Honório, D. A. Barbosa, E. J. Oliveira, P. A. N. Garcia, and M. F. Santos, "A multiple kernel classification approach based on a quadratic successive geometric segmentation methodology with a fault diagnosis case," *ISA Trans.*, vol. 74, pp. 209–216, Mar. 2018.
- [38] C. Zhang, X. Gao, Y. Li, and L. Feng, "Fault detection strategy based on weighted distance of k nearest neighbors for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 1, pp. 75–81, Jun. 2019.
- [39] Q. P. He and J. Wang, "Fault detection using the k -nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345–354, Nov. 2007.
- [40] Z. Zhou, C. Wen, and C. Yang, "Fault detection using random projections and k -Nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 70–79, Feb. 2015.
- [41] Y. Li and X. Zhang, "Diffusion maps based k -nearest-neighbor rule technique for semiconductor manufacturing process fault detection," *Chemometric Intell. Lab. Syst.*, vol. 136, pp. 47–57, Aug. 2014.
- [42] X. Hong, S. Chen, and C. J. Harris, "A forward-constrained regression algorithm for sparse kernel density estimation," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 193–198, Jan. 2008.
- [43] H. Zhang, X. Tian, X. Deng, and Y. Cao, "Batch process fault detection and identification based on discriminant global preserving kernel slow feature analysis," *ISA Trans.*, vol. 79, pp. 108–126, Aug. 2018.
- [44] H. Zhang, X. Tian, X. Deng, and Y. Cao, "Multiphase batch process with transitions monitoring based on global preserving statistics slow feature analysis," *Neurocomputing*, vol. 293, pp. 64–86, Jun. 2018.
- [45] N. Li and Y. Yang, "Ensemble kernel principal component analysis for improved nonlinear process monitoring," *Ind. Eng. Chem. Res.*, vol. 54, no. 1, pp. 318–329, Jan. 2015.
- [46] X. Deng and X. Tian, "Nonlinear process fault pattern recognition using statistics kernel PCA similarity factor," *Neurocomputing*, vol. 121, pp. 298–308, Dec. 2013.
- [47] F. J. Pontes, G. F. Amorim, P. P. Balestrassi, A. P. Paiva, and J. R. Ferreira, "Design of experiments and focused grid search for neural network parameter optimization," *Neurocomputing*, vol. 186, pp. 22–34, Apr. 2016.

- [48] W. Zong, G.-B. Huang, and Y. Chen, "Weighted extreme learning machine for imbalance learning," *Neurocomputing*, vol. 101, pp. 229–242, Feb. 2013.
- [49] J. Huang, X. Yang, and X. Yan, "Slow feature analysis-independent component analysis based integrated monitoring approach for industrial processes incorporating dynamic and static characteristics," *Control Eng. Pract.*, vol. 102, Sep. 2020, Art. no. 104558.
- [50] X. Ma, Y. Si, Z. Yuan, Y. Qin, and Y. Wang, "Multistep dynamic slow feature analysis for industrial process monitoring," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9535–9548, Dec. 2020, doi: 10.1109/TIM.2020.3004681.
- [51] K. Zhong, D. Ma, and M. Han, "Distributed dynamic process monitoring based on dynamic slow feature analysis with minimal redundancy maximal relevance," *Control Eng. Pract.*, vol. 104, Nov. 2020, Art. no. 104627.
- [52] L. Chiang, E. Russell, and R. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. London, U.K.: Springer-Verlag, 2001.



LANYUN SHAO received the B.E. degree from the School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan, China, and the master's degree from the School of Computer Science and Technology, Shandong University, Jinan. She is currently an Associate Professor with the School of Information and Electrical Engineering, Shandong Jianzhu University. Her major research interests include process modeling and simulation, fault diagnosis of industrial processes, and controller performance monitoring.



RONGBAO KANG is currently an In-Service Doctor of Engineering with the School of Cyberspace Security, University of Science and Technology of China. His main research interests include process modeling and monitoring as well as the analysis and verification of industrial control system security.



WEILIN YI received the B.E. and M.Sc. degrees from the College of Information and Control Engineering, China University of Petroleum (East China), Qingdao, China, in 2014 and 2017, respectively. He is currently an Engineer with The No. 30 Institute, CETC. His research interests include fault detection and diagnosis in industrial processes as well as cyberspace security of industrial control systems.



HANYUAN ZHANG received the B.E. degree from the School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan, China, in 2011, and the Ph.D. degree from the College of Information and Control Engineering, China University of Petroleum (East China), Qingdao, China, in 2017. He is currently a Lecturer with the School of Information and Electrical Engineering, Shandong Jianzhu University. His research interests include fault detection and diagnosis in industrial process as well as process modeling and simulation.

• • •