IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 December 2024, accepted 16 January 2025, date of publication 20 January 2025, date of current version 23 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3531712

RESEARCH ARTICLE

2DUMAP: Two-Dimensional Uniform Manifold Approximation and Projection for Fault Diagnosis

BENCHAO LI*, YUANYUAN ZHENG*, AND RUISHENG RAN, (Member, IEEE)
College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China

Corresponding author: Ruisheng Ran (rshran@cqnu.edu.cn)

This work was supported in part by Chongqing Municipal Education Commission under Grant KJZD-K202100505, in part by Chongqing Science and Technology Bureau under Grant cstc2020jscx-msxmX0190, and in part by the Ministry of Education of China under Grant 20YJAZH084.

*Benchao Li and Yuanyuan Zheng contributed equally to this work.

ABSTRACT With the continuous development of industry, the operational data of mechanical equipment has grown exponentially. High-dimensional fault data often contain a significant amount of noise signals and redundant information, severely impacting the performance of fault diagnosis methods. The currently popular manifold learning dimensionality reduction algorithm, Uniform Manifold Approximation and Projection (UMAP), effectively extract complex nonlinear relationships in fault diagnosis data while preserving both local and global structures. However, the UMAP has Out-of-Sample Problem, making it unable to directly map new samples to the learned low-dimensional space. To address this problem, a mapping matrix is learned to directly extend Out-of-Sample data, but this approach still faces the issue of the mapping matrix being too large. To tackle this challenge, the Two-Dimensional UMAP method is proposed, which transforms one-dimensional vector inputs into two-dimensional matrix representations. Subsequently, considering discriminant information, a Supervised Two-dimensional UMAP algorithm (2DSUMAP) is proposed. This study conducts fault diagnosis experiments on 7 datasets of bearings and gears to comprehensively evaluate the proposed methods. The experiments demonstrate that these methods solve the inherent Out-of-Sample Problem of the UMAP, effectively reduce algorithm complexity, and have superior classification performance and strong robustness in practical fault diagnosis applications.

INDEX TERMS UMAP, out-of-sample problem, fault diagnosis, manifold learning.

I. INTRODUCTION

With the continuous advancement of industrial development, the complexity of mechanical equipment has significantly increased, rendering mechanical failures a pivotal factor contributing to economic losses and posing threats to personnel safety [1], [2], [3]. Consequently, fault diagnosis has become particularly important. Fault diagnosis, which involves detecting, isolating, and identifying equipment faults [4], assists maintenance personnel in promptly identifying fault types or preventing faults, thereby reducing mechanical equipment downtime. Initially, fault diagnosis relied on on-site experiential judgments by maintenance personnel, requiring professional knowledge and incurring high time costs [1]. The advent of machine learning methods such as K-Nearest Neighbors (KNN) [5] and Support Vector Machine (SVM) [6] led to intelligent fault diagnosis, significantly improving fault identification efficiency.

However, with the development of sensors and computer systems, the dimensionality of data describing equipment operation has been continually increasing, exhibiting characteristics such as high dimensionality, strong coupling, and nonlinearity [7]. Additionally, these data contain a considerable amount of noise, significantly impacting the effectiveness of fault diagnosis. Therefore, there is an urgent need for dimensionality reduction methods to extract effective features from the data. Some researchers have used Principal Component Analysis (PCA) and its improved algorithms for feature extraction and dimensionality reduction [8], [9], [10]. However, PCA can only ensure that

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

the decomposed signal components are uncorrelated and cannot guarantee their independence. To overcome this limitation, Li et al. [11] introduced Independent Component Analysis (ICA) and proposed a new fault diagnosis method within the compressed sensing framework. Additionally, to preserve the local structure of the data, Wang et al. [12] combined Locality Preserving Projection (LPP) with PCA to propose a multi-modal process monitoring method based on Dynamic Locality Preserving Principal Component Analysis (DLPPCA), which addressed the modeling challenges caused by the strong dynamics of industrial processes. Meanwhile, Chen et al. [13] improved the Neighborhood Preserving Embedding (NPE) algorithm and proposed a scalable and secure Industrial Cyber-Physical System to enhance the effectiveness of early fault diagnosis.

Nevertheless, the aforementioned methods all belong to linear dimensionality reduction techniques, which have limited capability in handling nonlinear data. The emergence of manifold learning provides the possibility of extracting nonlinear features. McInnes et al. [14] proposed the Uniform Manifold Approximation and Projection (UMAP), which has significant advantages in preserving local structure, discovering nonlinear relationships, and handling large-scale data. However, its application in the field of fault diagnosis is currently limited. In this field, Joswiak et al. [15] investigated the advantages of the UMAP in chemical engineering fault diagnosis compared to other dimensionality reduction algorithms, aiming to enhance process understanding and fault troubleshooting capabilities in industrial environments. An et al. [16] utilized UMAP to explore more clustering manifold structures in the dataset, thereby providing better global structural information in latent space. Zhang and Shang [17] combined UMAP with metric learning to improve the accuracy of fault diagnosis.

Although the UMAP can handle more complex nonlinear relationships and better preserve local structure, it does not support directly mapping new samples to the low-dimensional space. This issue is known as the Out-of-Sample Problem in manifold learning [18], [19]. Currently, solutions to the Out-of-Sample Problem can be broadly categorized into three types [20]: linear framework-based methods [21], [22], [23], regression-based methods [24], [25] and kernel-based methods [26], [27]. Inspired by the aforementioned research, this paper proposes a method based on a linearization framework, aimed at addressing the Out-of-Sample Problem of UMAP and applying it to the field of intelligent fault diagnosis. This method is termed Linear Uniform Manifold Approximation and Projection (LUMAP).

The aforementioned dimensionality reduction methods, including PCA, LPP, NPE, and LUMAP, are primarily designed for vector data, necessitating the conversion of 2D data such as images into 1D format for application. When the original data has a high dimensionality, the size of the mapping matrix constructed by these methods becomes exceedingly large, thereby increasing the computational cost of the algorithms. Consequently, Yang et al. introduced the Two-Dimensional Principal Component Analysis (2DPCA) algorithm [28], which directly processes 2D data, effectively extracts key features from it, and reduces computational costs. Additionally, other dimensionality reduction algorithms tailored for 2D data include Two-Dimensional Linear Discriminant Analysis (2DLDA) [29] and Two-Dimensional Locality Preserving Projections (2DLPP) [30]. Yang et al. [28] demonstrated that matrix input enhances feature extraction speed and enables more accurate estimation of the mapping matrix. Consequently, researchers have proposed a series of fault diagnosis methods based on matrix inputs [31], [32], [33]. In this paper, we propose Two-Dimensional Uniform Manifold Approximation and Projection (2DUMAP) and Two-Dimensional Supervised Uniform Manifold Approximation and Projection (2DSUMAP) specifically tailored for 2D data, with the aims of reducing the computational cost of the algorithms and enhancing the accuracy of fault diagnosis.

The primary contributions of this paper are as follows:

- A linear dimensionality reduction method is proposed based on the linearization framework, which retains the ability of UMAP to extract nonlinear features from high-dimensional data while addressing the Out-of-Sample Problem inherent in the UMAP algorithm.
- Building upon Linear UMAP, 2DUMAP and 2DSUMAP are further introduced, significantly enhancing the feature extraction capability of the Linear UMAP method. Additionally, these methods effectively reduce the computational cost associated with the Linear UMAP approach.
- The 2DUMAP and 2DSUMAP methods proposed in this study are successfully applied to the field of intelligent fault diagnosis. Compared to existing machine learning methods, 2DSUMAP demonstrates superior recognition accuracy.

The remainder of this paper is organized as follows: In Section II, we review the related works of this study; in Section III, we elaborate on the principles of the proposed methods, including LUMAP, 2DUMAP, and 2DSUMAP; in Section IV, we evaluate the performance of the new methods; and in Section V, we present our conclusions.

II. RELATED WORKS

A. UMAP

Uniform Manifold Approximation and Projection (UMAP) represents a state-of-the-art methodology for dimensionality reduction and visualization, exhibiting extensive applicability across various domains such as bioinformatics [34], hyperspectral analysis [35], and environmental science [36]. In contrast to traditional dimensionality reduction techniques including PCA, Locally Linear Embedding (LLE)

[37], and t-Distributed Stochastic Neighbor Embedding (t-SNE) [38], UMAP demonstrates superior capability in uncovering the nonlinear characteristics of data while preserving the local structure among samples. Consequently, UMAP serves as a valuable adjunct in machine learning tasks such as pattern recognition and clustering analysis.

The procedure of UMAP can be summarized into two academically refined steps: Graph Construction and Graph Layout.

1) GRAPH CONSTRUCTION

For a given sample set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$, the set $\{\boldsymbol{x}_i^1, \boldsymbol{x}_i^2, \ldots, \boldsymbol{x}_i^k\}$ of k nearest neighbors for sample $\boldsymbol{x}_i$ can be readily obtained. Based on these nearest neighbors of the samples, a weighted directed graph $\overline{G} = (V, E, \omega)$ can be constructed, where $E = \{(\boldsymbol{x}_i, \boldsymbol{x}_i^j) | 1 \leq i \leq N, 1 \leq j \leq k\}$ denotes the set of edges in $\overline{G}$. The similarity ω between samples can be specifically expressed as:

$$\omega\left(\boldsymbol{x}_i, \boldsymbol{x}_i^j\right) = \exp\left(\frac{-\max(0, d(\boldsymbol{x}_i, \boldsymbol{x}_i^j) - \rho_i)}{\sigma_i}\right) \tag{1}$$

where, ρ_i denotes the distance between the sample $\boldsymbol{x}_i$ and its nearest neighbor, while σ_i represents the normalization factor for the sample $\boldsymbol{x}_i$, which can be derived or solved through $\Sigma_{j=1}^{k} \omega_{ij} = \log_2 k$.

The similarity between samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be computed via $\omega_{i|j}$ and $\omega_{j|i}$. Given the adjacency matrix $\boldsymbol{A}$ of the directed weighted graph $\overline{G}$, the adjacency matrix $\boldsymbol{B}$ of corresponding undirected weighted graph G can be represented as follows:

$$\boldsymbol{B} = \boldsymbol{A} + \boldsymbol{A}^T - \boldsymbol{A} \circ \boldsymbol{A}^T \tag{2}$$

2) GRAPH LAYOUT

In the low-dimensional space, the similarity υ_{ij} between the projected points $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ of the samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be expressed as:

$$\upsilon\left(\boldsymbol{y}_i, \boldsymbol{y}_j\right) = \left(1 + a\left(\left\|\boldsymbol{y}_i - \boldsymbol{y}_j\right\|_2^2\right)^b\right)^{-1} \tag{3}$$

where, a and b represent two hyperparameters. UMAP continuously optimizes the coordinates of the projected points on the low-dimensional manifold by minimizing the cross-entropy function 4. This process aims to approximate the distribution of projected points on the low-dimensional manifold to the distribution of samples on the high-dimensional manifold, thereby achieving the purpose of dimensionality reduction.

$$\mathrm{CE} = \sum \omega \log\left(\frac{\omega}{\upsilon}\right) + (1 - \omega) \log\left(\frac{1 - \omega}{1 - \upsilon}\right) \tag{4}$$

Specifically, UMAP optimizes its embeddings $\boldsymbol{y}_i$ through iterative application of attractive force $\mathcal{F}_{UMAP} = \nabla \omega \log \upsilon$ and repulsive force $\mathcal{H}_{UMAP} = \nabla(1 - \omega) \log(1 - \upsilon)$ acting upon the points in the projected space.

B. LINEAR METHODS

The nonlinear manifold learning techniques, including LLE, Laplacian Eigenmaps (LE) [39], t-SNE, and UMAP, are all plagued by the Out-of-Sample Problem, which restricts their capability to project unknown new samples into the low-dimensional space. To address the Out-of-Sample Problem associated with LLE, LE, and t-SNE, a series of methods based on the concept of linearization have been proposed, such as LPP, NPE, and Linear t-SNE (Lt-SNE) [23]. These linear methods all adhere to the general form outlined in Equation 5, utilizing a mapping matrix $\boldsymbol{A}$ to project samples from the high-dimensional space into the low-dimensional space.

$$\boldsymbol{Y} = \boldsymbol{A}^T \boldsymbol{X} \tag{5}$$

Additionally, a range of dimensionality reduction techniques tailored for 2D data, including 2DPCA [28], 2DLDA [29], and 2DLPP [30], are also linear methods that satisfy the general form in Equation 5.

III. METHODS

A. LUMAP

The objective of LUMAP is to learn a mapping matrix $\boldsymbol{A}$ that projects samples $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$ on the high-dimensional manifold onto the low-dimensional manifold through the general form of linear methods as expressed in Equation 5. To approximate the topological structure of the projected points $\{\boldsymbol{A}^T\boldsymbol{x}_1, \boldsymbol{A}^T\boldsymbol{x}_2, \ldots, \boldsymbol{A}^T\boldsymbol{x}_N\}$ on the low-dimensional manifold to that of the samples on the high-dimensional manifold, LUMAP continuously optimizes the mapping matrix $\boldsymbol{A}$ by iteratively applying attractive force $\mathcal{F}_{LUMAP}$ and repulsive force $\mathcal{H}_{LUMAP}$ between projected points $\boldsymbol{A}^T\boldsymbol{x}_i$ and $\boldsymbol{A}^T\boldsymbol{x}_j$.

In the LUMAP method, the computation of similarity ω_{ij} between samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ on the high-dimensional manifold is consistent with the computation method employed in the UMAP method. By combining the similarity between projected points on the low-dimensional manifold in the UMAP method, as described in Section II-A, with the general form 5 of linear dimensionality reduction methods outlined in Section II-B, the similarity υ_{ij} between projected points $\boldsymbol{A}^T\boldsymbol{x}_i$ and $\boldsymbol{A}^T\boldsymbol{x}_j$ on the low-dimensional manifold in the LUMAP method can be defined as:

$$\upsilon(\boldsymbol{A}^T\boldsymbol{x}_i, \boldsymbol{A}^T\boldsymbol{x}_j) = \left(1 + a\left(\left\|\boldsymbol{A}^T\boldsymbol{x}_i - \boldsymbol{A}^T\boldsymbol{x}_j\right\|_2^2\right)^b\right)^{-1} \tag{6}$$

Similar to UMAP, LUMAP optimizes the mapping matrix $\boldsymbol{A}$ by minimizing the cross-entropy function 4. The attractive force $\mathcal{F}_{LUMAP}$ and repulsive force $\mathcal{H}_{LUMAP}$ between projected points on the low-dimensional manifold can be specifically

defined as:

$$\mathcal{F}_{LUMAP} = \frac{\partial \omega \log \upsilon}{\partial \boldsymbol{A}} \tag{7}$$

$$= \frac{-2ab\omega\left(\left\|\Delta_{LUMAP}\right\|_2^2\right)^{b-1}\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)\Delta_{LUMAP}^T}{1 + a\left(\left\|\Delta_{LUMAP}\right\|_2^2\right)^b} \tag{8}$$

$$\mathcal{H}_{LUMAP} = \frac{\partial\left(1-\omega\right)\log\left(1-\upsilon\right)}{\partial \boldsymbol{A}} \tag{9}$$

$$= \frac{2b\left(1-\omega\right)\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)\Delta_{LUMAP}^T}{\left[\epsilon + \left\|\Delta_{LUMAP}\right\|_2^2\right]\left[1 + a\left(\left\|\Delta_{LUMAP}\right\|_2^2\right)^b\right]} \tag{10}$$

where, $\Delta_{LUMAP} = \boldsymbol{A}^T\boldsymbol{x}_i - \boldsymbol{A}^T\boldsymbol{x}_j$. Additionally, to avoid the scenario where $\|\Delta_{LUMAP}\|_2 = 0$, we introduce a small positive number ϵ as a regularization term in Equation 10. In experiments, ϵ is typically set to 0.001.

B. 2DUMAP

The advantages of dimensionality reduction methods tailored for 2D data have been demonstrated in [28]. Inspired by methods like 2DPCA, 2DLDA, and 2DLPP, we propose 2DUMAP based on the general form of linear dimensionality reduction methods outlined in Equation 5. 2DUMAP achieves dimensionality reduction of 2D sample $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ through the mapping matrix $\boldsymbol{A} \in \mathbb{R}^{n\times d}$, thereby obtaining the critical features $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A}, \boldsymbol{Y} \in \mathbb{R}^{m\times d}$.

For a sample set $\mathcal{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N\}$ residing on the high-dimensional manifold, 2DUMAP iteratively optimizes the mapping matrix $\boldsymbol{A}$ to obtain a set of projection points $\mathcal{Y} = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_N\}$ on the low-dimensional manifold. To minimize the discrepancy between the topological structure of the sample set $\mathcal{X}$ and the topological structure of the projection points $\mathcal{Y}$, 2DUMAP redefines the similarity between projection points on the low-dimensional manifold and achieves dimensionality reduction by minimizing a cross-entropy function 4. The similarity between projection points on the low-dimensional manifold is defined as:

$$\upsilon(\boldsymbol{X}_i\boldsymbol{A}, \boldsymbol{X}_j\boldsymbol{A}) = \left[1 + a\left(\left\|\boldsymbol{X}_i\boldsymbol{A} - \boldsymbol{X}_j\boldsymbol{A}\right\|_F^2\right)^b\right]^{-1} \tag{11}$$

Similar to LUMAP, 2DUMAP applies attractive force $\mathcal{F}_{2DUMAP}$ and repulsive force $\mathcal{H}_{2DUMAP}$ between the projection points on the low-dimensional manifold during the optimization process. The attractive force $\mathcal{F}_{2DUMAP}$ and repulsive force $\mathcal{H}_{2DUMAP}$ exerted by 2DUMAP on the projection points during optimization are respectively defined as:

$$\mathcal{F}_{2DUMAP} = \frac{-2ab\omega\left(\left\|\Delta_{2DUMAP}\right\|_F^2\right)^{b-1}\left(\boldsymbol{X}_i^T - \boldsymbol{X}_j^T\right)\Delta_{2DUMAP}}{1 + a\left(\left\|\Delta_{2DUMAP}\right\|_F^2\right)^b} \tag{12}$$

$$\mathcal{H}_{2DUMAP} = \frac{2b\left(1-\omega\right)\left(\boldsymbol{X}_i^T - \boldsymbol{X}_j^T\right)\Delta_{2DUMAP}}{\left[\epsilon + \left\|\Delta_{2DUMAP}\right\|_F^2\right]\left[1 + a\left(\left\|\Delta_{2DUMAP}\right\|_F^2\right)^b\right]} \tag{13}$$

where, $\Delta_{2DUMAP} = \boldsymbol{X}_i\boldsymbol{A} - \boldsymbol{X}_j\boldsymbol{A}$. To avoid the situation where $\|\Delta_{2DUMAP}\|_F = 0$ in Equation 13, we introduce a small positive number ϵ, which is typically set to 0.001 in experiments.

C. 2DSUMAP

The LUMAP proposed in Section III-A, the 2DUMAP introduced in Section III-B, and the original UMAP method are all unsupervised learning approaches that fail to fully exploit the label information of samples. To fully leverage the label information of the data, better preserve the manifold structure of samples, and enhance the discriminative capability of the mapping matrix, this section proposes 2DSUMAP.

Due to the difficulty of accurately describing the positional relationships between categories using 1-dimensional labels in Euclidean space, this study redefines the label of samples through NPE technology. NPE is employed to embed high-dimensional samples onto the low-dimensional manifold, and the coordinates of the central points of each category on this low-dimensional manifold are taken as the new labels for that category. For a sample set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$ belonging to C categories, the specific form of the new labels can be defined as:

$$\boldsymbol{t}_i = \frac{1}{N_i}\sum_{j}^{N_i}\mathrm{NPE}\left(\boldsymbol{x}_i^j, \eta\right) \quad i = 1, 2, 3, \ldots, C \tag{14}$$

where, N_i denotes the number of samples in the i-th class, and $\boldsymbol{x}_i^j$ represents the j-th sample within the i-th class. In 2DUMAP, the distance between samples is redefined as:

$$d_s\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right) = \lambda \cdot \left\|\boldsymbol{X}_i - \boldsymbol{X}_j\right\| + (1-\lambda)\cdot\left\|\boldsymbol{t}_i - \boldsymbol{t}_j\right\| \tag{15}$$

In this context, $\lambda(\lambda \in (0, 1))$ and $1-\lambda$ represent the respective proportions of data information and label information that contribute to the measurement of sample distances.

By applying the newly defined metric d_s to the graph construction of LUMAP and 2DUMAP respectively, the corresponding Supervised Linear Uniform Manifold Approximation and Projection (SLUMAP) and Two-Dimensional Supervised Uniform Manifold Approximation and Projection (2DSUMAP) can be proposed.

D. 2DUMAP FOR FAULT DIAGNOSIS

Figure 1 illustrates the feature extraction and classification processes of LUMAP, 2DUMAP, and 2DSUMAP for mechanical vibration signals. The sequence from (a) to (c) to (d) depicts the feature extraction process using LUMAP for vibration signals. Initially, a neighborhood graph is constructed on the high-dimensional Riemannian manifold in (c), followed by the optimization of projection points on the low-dimensional Riemannian manifold in (d). The

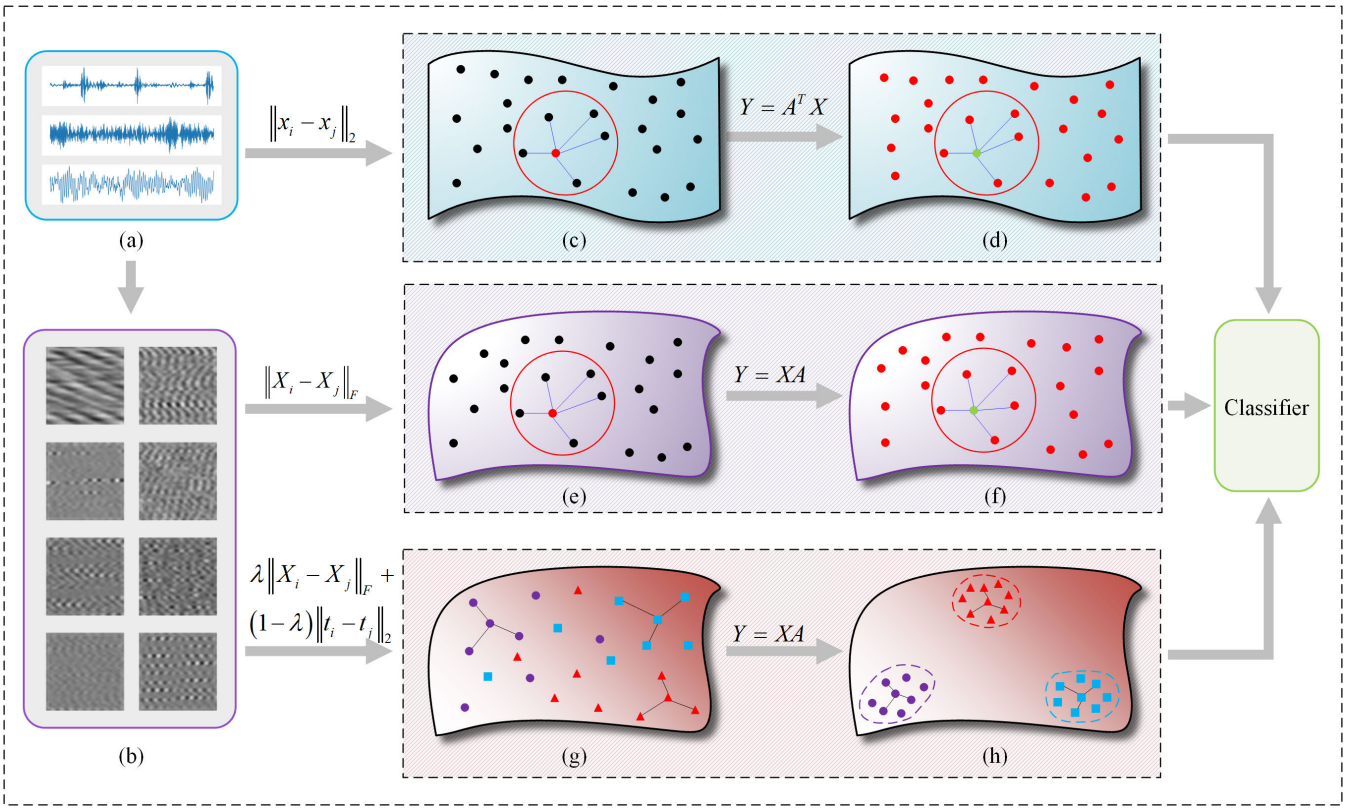


FIGURE 1. Fundamental concepts of LUMAP, 2DUMAP, and 2DSUMAP for fault diagnosis: (a) 1D Vibration signals; (b) Resampled 2D vibration signals; (c) Neighborhood graph of 1D vibration signals constructed by LUMAP; (d) Key features of vibration signals extracted by LUMAP; (e) Neighborhood graph of 2D vibration signals constructed by 2DUMAP; (f) Key features of vibration signals extracted by 2DUMAP; (g) Neighborhood graph of 2D Vibration Signals Constructed by 2DSUMAP; (h) Key features of vibration signals extracted by 2DSUMAP.

sequence from (a) to (b) to (e) to (f) represents the feature extraction process using 2DUMAP for 2D vibration signals. The topological structure of the low-dimensional manifold in (f) is optimized by applying attractive and repulsive forces to the projection points residing on it. The sequence from (a) to (b) to (g) to (h) denotes the feature extraction process using 2DSUMAP for 2D vibration signals. By redefining the distances between samples in (g), the distances between projection points of the same class on the low-dimensional manifold in (h) are minimized, while those between projection points of different classes are maximized.

IV. EXPERIMENTS

To demonstrate the advantages of the 2DUMAP and 2DSUMAP methods in fault diagnosis, we have carried out the following tasks:

(1) We analyzed the optimal target dimension and training data ratio across multiple fault diagnosis datasets, and compared their runtime performance;
(2) We compared the effectiveness of the proposed 2DUMAP and 2DSUMAP methods with other classical dimensionality reduction techniques on multiple fault diagnosis datasets, using two classifiers: KNN and SVM.

A. DATASETS

In this study, we employed 3 bearing datasets and 2 gear datasets to evaluate the effectiveness of the 2DUMAP and 2DSUMAP methods. The key technical parameters of these datasets are summarized in Table 1.

The Bearing Dataset from the University of Ottawa [40] contains vibration signals and rotational speed data for 3 conditions: healthy, inner ring fault, and outer ring fault, each recorded under 4 different rotational speeds. In this study, vibration signals from the healthy condition, inner ring fault, and outer ring fault under accelerated conditions were selected. Each sample was then processed and resized into a 2D image of 32 × 32 pixels for analysis.

The bearing dataset of Politecnico di Torino [41] encompasses vibration signals from inner ring faults and roller faults at sizes of 150, 250, and 450 μm, as well as signals from normal bearings. In this study, the vibration signals from inner ring faults and roller faults at 250 and 450 μm, along with signals from healthy bearings, were selected to assess the validity of the proposed testing model. In experiments, each sample was resized to a 2D image of 32 × 32 pixels.

The Southeast University Bearing [42] Dataset comprises vibration signals collected under 2 operational conditions: 20Hz-0V and 30Hz-2V. These signals encompass data from

TABLE 1. Key technical parameters of the dataset.

Datasets	Abbreviation	Type	Sample Number/Class	Class Number	Sample Size
University of Ottawa	B1	Bearing	300	3	32×32
Politecnico di Torino	B2	Bearing	118	5	32×32
Southeast University	B3	Bearing	500	5	32×32
University of Connecticut	G1	Gear	104	9	32×32
Southeast University	G2	Gear	500	5	32×32
Mixed Bearing	MB	Bearing	500	5	32×32
Mixed Gear	MG	Gear	100	12	32×32

TABLE 2. Composition of the mixed bearing dataset.

Datasets	B1		B2		B3		MB
	Total	Sampled	Total	Sampled	Total	Sampled	
Health	300	164	118	64	500	272	500
Inner	300	145	236	114	500	241	500
Outer	300	188	-	-	500	312	500
Roller/Ball	-	-	236	160	500	340	500
Combination	-	-	-	-	500	500	500

TABLE 3. Composition of the mixed gear dataset.

Datasets	G1		G2		MG
	Total	Sampled	Total	Sampled	
Health	104	50	500	50	100
Missing	104	50	500	50	100
Chipped	-	-	500	100	100
Root	-	-	500	100	100
Surface	-	-	500	100	100
Chip1a	104	100	-	-	100
Chip2a	104	100	-	-	100
Chip3a	104	100	-	-	100
Chip4a	104	100	-	-	100
Chip5a	104	100	-	-	100
Spall	104	100	-	-	100
Crack	104	100	-	-	100

5 health conditions, namely ball fault, inner ring fault, outer ring fault, combination fault on both inner and outer rings, and normal working state. During the data preprocessing phase, each sample is resampled into a 2D image of 32×32 pixels for subsequent analysis.

The Gear Dataset from the University of Connecticut [43] encompasses vibration signals from 9 distinct health conditions: healthy, missing, crack, spall, as well as chipped conditions labeled as chip1a, chip2a, chip3a, chip4a, and chip5a. Each signal has been transformed and resized into a 2D image of dimensions 32×32 pixels for analysis.

The Southeast University Gear [42] Dataset encompasses eight distinct vibration signals, which encompass data from 5 different conditions: chipped tooth, missing tooth, root fault, surface fault, and normal working state. Each sample within the dataset has been resized to a 2D image of 32×32 pixels for analytical purposes.

To evaluate the effectiveness and stability of 2DUMAP and 2DSUMAP in more complex environments, we mixed 3 bearing datasets and 2 gear datasets into 2 novel consolidated datasets, respectively. The 3 bearing datasets collectively encompass 5 modes: inner ring fault, outer ring fault, roller fault, combination fault, and normal condition. We employed stratified sampling to extract samples from these 3 datasets to synthesize a new hybrid dataset. The 2 gear datasets, on the other hand, encompass 12 states: normal, missing, crack, spall, surface, root, chipped, chip1a, chip2a, chip3a, chip4a, and chip5a, random sampling was utilized to select samples to compile a new consolidated dataset.

The key technical parameters of the 2 mixed datasets are documented in Table 1. The composition of various fault samples within the MB and MG datasets is detailed in Tables 2 and 3, respectively.

B. PERFORMANCE COMPARISON

This section presents the testing of the 2DUMAP and 2DSUMAP algorithms on datasets B1, B2, B3, G1, G2, MB, and MG to evaluate their performance in fault diagnosis tasks. Since both 2DUMAP and 2DSUMAP are linear dimensionality reduction methods, we have conducted comparisons with classic dimensionality reduction algorithms such as PCA, ICA, LPP, Probabilistic Nearest Neighbors Based Locality Preserving Projections (PNNLPP) [44], NPE, UMAP, Parametric UMAP (PUMAP) [45], and others. Additionally, we have also compared their performance with 2DPCA.

In this study, we utilized 20% of the samples as training data and the remaining 80% as testing data. Tables 4 and 5 present the experimental results of the features extracted by the aforementioned methods on the KNN classifier and SVM classifier, respectively.

Compared to traditional linear dimensionality reduction methods such as PCA, LPP, and NPE, the classification results of 2DUMAP and 2DSUMAP exhibit significant improvements. Furthermore, the performance of 2DUMAP

TABLE 4. Performance comparison of various algorithms on the KNN classifier.

Method	B1	B2	B3	G1	G2	MB	MG
PCA	0.9569	0.4006	0.7981	0.7047	0.7091	0.7732	0.6961
2DPCA	0.9760	0.7734	0.8181	0.5582	1.0000	0.7935	0.7318
ICA	0.4514	0.3955	0.5324	0.6649	0.4017	0.6191	0.5103
LPP	0.3694	0.4136	0.3927	0.7182	0.3180	0.4099	0.4911
PNNLPP	0.5356	0.4504	0.8049	0.5877	0.6725	0.7981	0.6506
NPE	0.3488	0.2389	0.2191	0.7360	0.2712	0.2649	0.4816
UMAP	0.9593	0.4791	0.8890	0.8195	0.7499	0.8532	0.7465
PUMAP	0.9461	0.6004	0.6773	0.6430	0.7182	0.6321	0.6235
2DUMAP	**0.9693**	**0.7870**	**0.8321**	**0.9269**	**1.0000**	**0.7955**	**0.8513**
2DSUMAP	**0.9661**	**0.9209**	**0.8359**	**0.9663**	**1.0000**	**0.7542**	**0.8700**

TABLE 5. Performance comparison of various algorithms on the SVM classifier.

Method	B1	B2	B3	G1	G2	MB	MG
PCA	0.9621	0.5936	0.6343	0.6953	0.9980	0.6493	0.6359
2DPCA	0.9667	0.7009	0.6325	0.4486	0.9976	0.6276	0.6071
ICA	0.7161	0.4068	0.6550	0.8329	0.7031	0.6030	0.6286
LPP	0.3497	0.4872	0.3153	0.6594	0.2387	0.3079	0.3964
PNNLPP	0.5153	0.3770	0.2978	0.5112	0.5232	0.6252	0.5298
NPE	0.3264	0.2538	0.2311	0.7655	0.2257	0.2822	0.4520
UMAP	0.9631	0.4013	0.8452	0.7452	0.6190	0.7811	0.6457
PUMAP	0.9464	0.5638	0.5771	0.5378	0.7865	0.5813	0.5101
2DUMAP	**0.9697**	**0.7891**	**0.6380**	**0.8415**	**1.0000**	**0.6433**	**0.6928**
2DSUMAP	**0.9679**	**0.8615**	**0.6201**	**0.8703**	**1.0000**	**0.5995**	**0.7114**

and 2DSUMAP surpasses that of PNNLPP, a novel linear dimensionality reduction method based on LPP. Both 2DPCA, 2DUMAP, and 2DSUMAP are linear dimensionality reduction techniques tailored for 2D data. In comparison to 2DPCA, 2DUMAP and 2DSUMAP achieve higher classification accuracy. Although both UMAP and PUMAP are nonlinear dimensionality reduction methods, 2DUMAP demonstrates superior classification performance across most datasets. On numerous datasets, 2DSUMAP outperforms 2DUMAP in fault identification, underscoring the importance of label information in model training.

To more fully demonstrate the advantages of 2DSUMAP, we have also conducted comparisons with some of the latest fault diagnosis methods. It is noteworthy that CNN-2560-768 [46], LeNet-5 [46], [47], and LiNet [48] are neural networks based on 2D data, whereas TICNN [49] and WDCNN [50] are neural networks tailored for vector data. These methods all belong to the category of supervised deep learning, adhering to the structural configurations outlined in their original research. The number of iterations for each method has been set to 100, ensuring consistency in the comparative analysis.

The recognition results of the aforementioned methods across various datasets are documented in Table 6. From the results presented in Table 6, it can be observed that 2DSUMAP achieves the highest classification accuracy on datasets such as B1, B2, B3, G2, and MG. For the MB and G1 dataset, 2DSUMAP also attains a relatively high classification accuracy. Overall, 2DSUMAP demonstrates satisfactory performance in fault diagnosis tasks.

C. 2DUMAP AND 2DSUMAP FOR OUT-OF-SAMPLE PROBLEM

In this section, we delve into the capability of 2DUMAP and 2DSUMAP in addressing the Out-of-Sample Problem. Specifically, both 2DUMAP and 2DSUMAP learn a mapping matrix $\boldsymbol{A}$ from the training data and utilize this matrix $\boldsymbol{A}$ to perform dimensionality reduction on Out-of-Sample data, thereby effectively resolving the Out-of-Sample Problem inherent in UMAP.

Assuming that the distribution of training data $\mathcal{X}_{tr}$ and Out-of-Sample data $\mathcal{X}_{oos}$ in the high-dimensional space is consistent, the distribution of training samples $\mathcal{Y}_{tr}$ in the low-dimensional space should align with the distribution of the projected points $\mathcal{Y}_{oos}$ obtained through the mapping matrix $\boldsymbol{A}$. In this study, we employed training samples $\mathcal{Y}_{tr}$ in the low-dimensional space to train both a KNN classifier and a SVM classifier. Subsequently, we utilized these trained classifiers to classify the Out-of-Sample points $\mathcal{Y}_{oos}$ in the low-dimensional space, obtained which through were the mapping matrix $\boldsymbol{A}$.

The higher the classification accuracy of the classifier, the closer the distribution of training samples and Out-of-Sample data in the low-dimensional space. Tables 4 and 5 document the classification results of the KNN classifier and SVM classifier, respectively. These results not only reflect the performance of 2DUMAP and 2DSUMAP in fault diagnosis tasks but also demonstrate their capability in addressing the Out-of-Sample Problem.

Currently, PUMAP offers a solution to the Out-of-Sample Problem inherent in UMAP. According to the results

TABLE 6. Performance comparison of various deep net algorithms with the KNN classifier.

Method	B1	B2	B3	G1	G2	MB	MG
CNN-2560-768	0.7625	0.7638	0.7925	0.9973	0.9995	0.8190	0.8764
LeNet-5	0.9694	0.7809	0.7005	0.9987	0.9205	0.7231	0.8520
LiNet	0.3333	0.2000	0.2695	0.1740	0.2000	0.2918	0.0985
TICNN	0.9625	0.8340	0.6895	0.9665	0.8995	0.6809	0.8244
WDCNN	0.9653	0.7404	0.7030	0.9237	0.9955	0.6488	0.8106
2DSUMAP-KNN	**0.9661**	**0.9209**	**0.8359**	**0.9663**	**1.0000**	**0.7542**	**0.8700**

presented in Tables 4 and 5, 2DSUMAP and 2DSUMAP demonstrate superior capability in addressing the Out-of-Sample Problem compared to PUMAP. When compared to classical UMAP, 2DUMAP and 2DSUMAP exhibit better dimensionality reduction performance for Out-of-Sample data. Additionally, we have compared the classification abilities of 2DUMAP and 2DSUMAP for Out-of-Sample points in low-dimensional spaces with those of classical feature extraction methods such as PCA, 2DPCA, and ICA. Based on the results in Tables 4 and 5, 2DSUMAP and 2DSUMAP surpass PCA, 2DPCA, ICA, and other traditional feature extraction methods in terms of feature extraction capability on most datasets.

D. PARAMETERS ANALYSIS

In this section, an analysis is conducted on the target dimensionality for dimensionality reduction and the proportion of training data utilized during the feature extraction process for both 2DUMAP and 2DSUMAP. Given the necessity of conducting multiple repetitive experiments in the analysis, two smaller-scale datasets, namely B1 and G1, are selected to economize computational costs.

The target dimensionality of dimensionality reduction methods can be determined through the Akaike Information Criterion (AIC) [51]. The specific formulation of AIC can be expressed as:

$$d = \text{argmin}\left[f_e(d^*) + \frac{d^*}{N}\right] \tag{16}$$

where, $f_e(d^*)$ represents the classification error rate of the extracted features when the dimensionality is d^*, N denotes the maximum dimensionality to be observed, and d signifies the optimal dimension in dimensionality reduction process.

We analyze the optimal target dimensionality for 2DUMAP and 2DSUMAP within the range of $[1, 30]$. The analytical results for the optimal target dimensionality of 2DUMAP and 2DSUMAP on the B1 and G1 datasets are reported in Figure 2.

An intuitive conclusion can be drawn from Figure 2, as the target dimensionality increases, the AIC value also increases continually. In this study, the target dimensionality for both 2DUMAP and 2DSUMAP across the various datasets has been determined to be 32×1.

We conducted an analysis of the optimal proportion of training data within the range of $[0.05, 0.95]$. Figure 3 presents the classification accuracy of 2DUMAP and 2DSUMAP under varying proportions of training data.

Examining the results for the B1 dataset in Figure 3(a), the impact of the training data proportion on classification accuracy is not pronounced. From the outcomes depicted in Figure 3(b), it is evident that when the proportion of training data reaches 20%, the classification accuracy of features extracted by both 2DUMAP and 2DSUMAP tends to stabilize. By integrating the findings from both Figure 3(a) and 3(b), in this study, we utilized the training data proportion of 20% for both 2DUMAP and 2DSUMAP models, and reserved 80% of the data for testing purposes.

E. COMPARISON OF RUNTIME

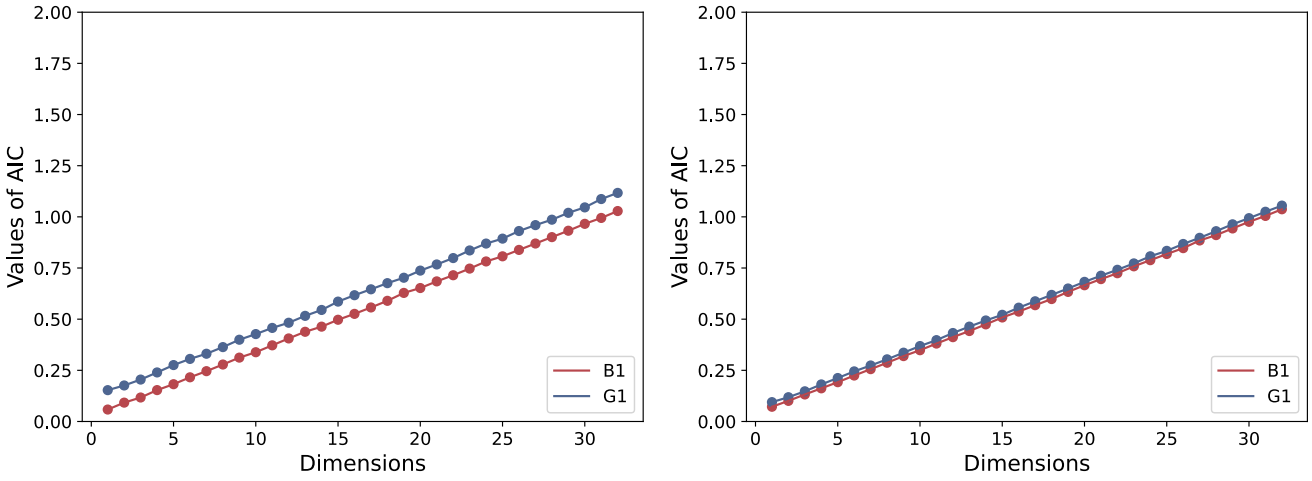
We have compared the runtime of UMAP, PUMAP, 2DUMAP, and 2DSUMAP across various datasets. Each of these algorithms was run 10 times on each dataset, and the average time cost is recorded in Table 7.

Both 2DUMAP and 2DSUMAP, along with PUMAP, can address the Out-of-Sample Problem of UMAP. However, compared to PUMAP, 2DUMAP, and 2DSUMAP achieve superior performance in a shorter period of time.

For a 1D sample set $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}, \boldsymbol{x}_i \in \mathbb{R}^{mn}$, the time complexity of the graph optimization phase in the classical UMAP algorithm is $O(KNmn)$. In contrast to the classical UMAP, both 2DUMAP, and 2DSUMAP necessitate multiple computations of $\boldsymbol{X}_i\boldsymbol{A}$ and $\boldsymbol{X}_j\boldsymbol{A}$ during each iteration. The computational cost for a single calculation of the inner product between the mapping matrix $\boldsymbol{A} \in \mathbb{R}^{n\times d}$ and a 2D sample $\boldsymbol{X}_i \in \mathbb{R}^{m\times n}$ is $O(mnd)$. Consequently, the additional time complexity introduced in the optimization process of 2DUMAP and 2DSUMAP, compared to the classical UMAP, is $O(kNmnd)$.

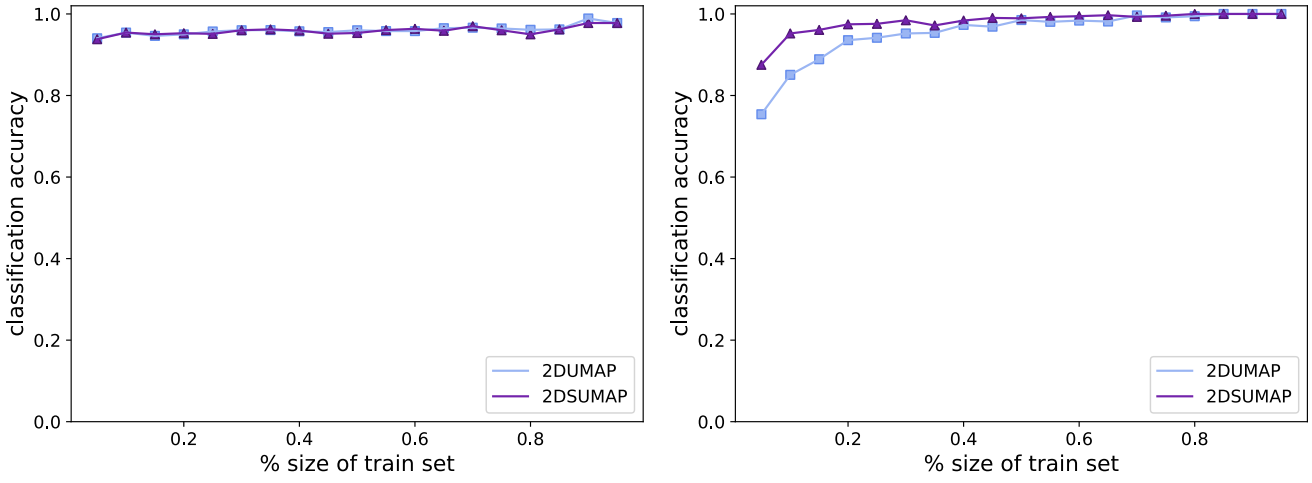
During the graph construction phases of UMAP and 2DUMAP, in order to compute the k-nearest neighbor graph for the samples, it is necessary to calculate the pairwise distance matrix for the samples, which entails a time complexity of $O(mnN^2)$. In the graph construction phase of 2DSUMAP, an additional computation is required for the pairwise distance matrix of the label vectors, thereby increasing the overall time complexity to $O(mnN^2) + O(\eta N^2)$, where η represents the dimensionality of the label vectors.

Based on the results presented in Table 7, the time cost of 2DUMAP is $1.5 - 3$ times higher than that of the classical UMAP, and the time cost of 2DSUMAP is slightly higher than



(a) 2DUMAP (b) 2DSUMAP

FIGURE 2. AIC values of 2DUMAP and 2DSUMAP on B1 and G1 datasets.



(a) B1 (b) G1

FIGURE 3. Classification accuracy of 2DUMAP and 2DSUMAP under various proportions of training data.

TABLE 7. Runtime(s) of UMAP, PUMAP, 2DUMAP, and 2DSUMAP on various datasets.

Method	B1	B2	B3	G1	G2	MB	MG
UMAP	11.0112	3.1588	8.9059	3.8657	9.0206	14.8318	10.4922
PUMAP	26.3492	16.8175	34.4879	17.7091	37.4119	40.9076	26.8105
2DUMAP	**16.1144**	**7.0635**	**14.2427**	8.**9658**	**15.9686**	**22.9851**	**17.3287**
2DSUMAP	**19.1906**	**9.7071**	**18.6617**	**8.9116**	**29.7708**	**27.9046**	**20.9412**

2DUMAP. This aligns with the aforementioned theoretical analysis.

The time cost of both 2DUMAP and 2DSUMAP can be effectively controlled by regulating the number of iterations in the graph optimization phase. Furthermore, employing acceleration tools such as Numba and leveraging high-performance computing devices like GPUs can further reduce the time cost associated with 2DUMAP and 2DSUMAP.

F. DISCUSSION ON MISCLASSIFICATIONS

We aim to analyze the scenarios where 2DUMAP fails to correctly identify fault samples. To this end, we conduct a detailed examination of the classification process of

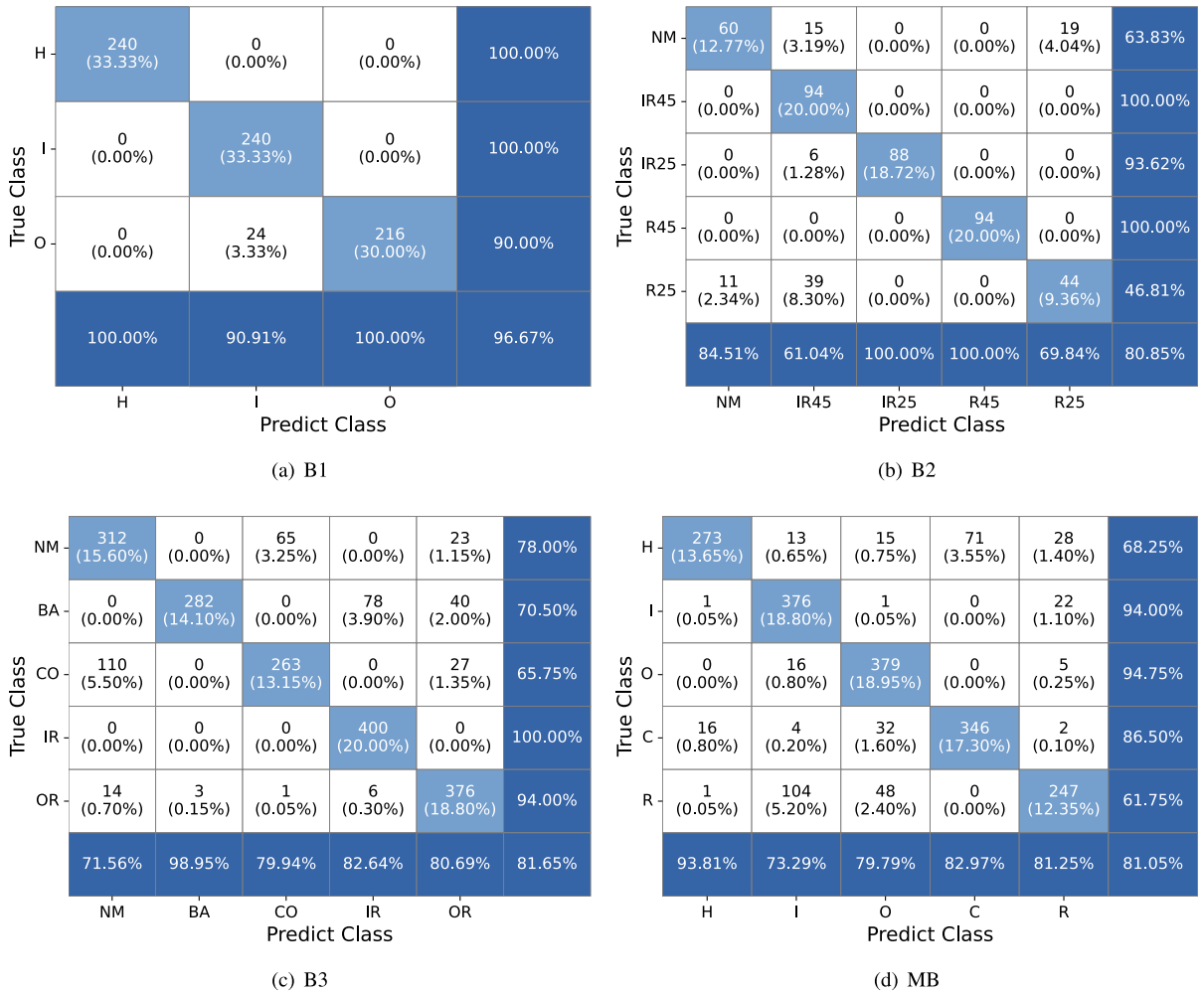


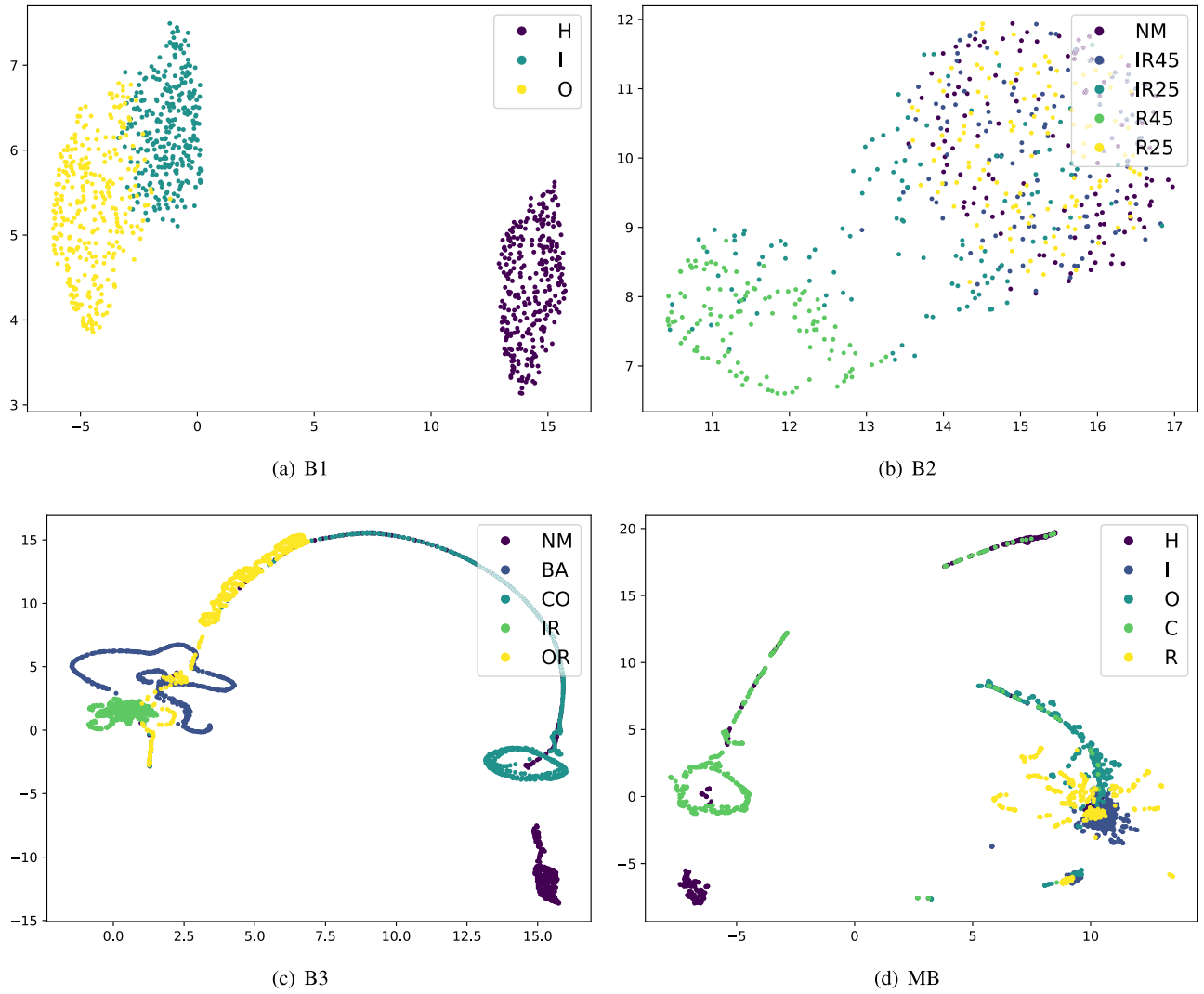
FIGURE 4. Confusion matrices representing the recognition results of the 2DUMAP-KNN Model for datasets B1, B2, B3, and MB.

2DUMAP using datasets such as B1, B2, B3, and MB as examples. The confusion matrices for the fault diagnosis results of the 2DUMAP-KNN model on these datasets are depicted in Figure 4.

From the results presented in Figure 4, it is evident that the 2DUMAP-KNN model demonstrates the best recognition performance for the B1 dataset, whereas its performance is relatively lower for the B2, B3, and MB datasets. In conjunction with the outcomes detailed in Tables 4 and 5, classical methods such as PCA, 2DPCA, and UMAP exhibit similar conclusions to 2DUMAP. To this end, we have utilized UMAP in Figure 5 to visualize the distribution of samples within the B1, B2, B3, and MB datasets, thereby analyzing the impact of sample distribution on recognition results.

Figure 4(a) presents the confusion matrix for the recognition results of the 2DUMAP-KNN model on the B1 dataset, where a small number of outer ring faults (O) are misclassified as inner ring faults (I). Figure 5(a) depicts the distribution of samples within the B1 dataset. In Figure 5(a), the distinct boundaries between various sample classes provide an explanation for the superior recognition performance of various methods on the B1 dataset. A small number of outer ring fault samples are interspersed within the inner ring fault sample cluster on the left side of Figure 5(a), indicating a stronger similarity between these outer ring fault samples and the inner ring fault samples. This explains the reason why the 2DUMAP-KNN model fails to correctly identify these specific outer ring fault samples.

From the confusion matrix of the B3 dataset presented in Figure 4(c), it is evident that the recognition accuracy for inner ring fault (IR) samples is 100%. Conversely, the recognition rates for normal (NM) samples, ball fault (BA) samples, and combination fault (CO) samples are relatively low. Specifically, some normal samples are misclassified as combination fault samples, and vice versa. This misclassification is attributed to the overlap of normal samples and combination fault samples observed in the right and upper sections of Figure 5(c). Additionally, the indistinct boundaries among ball fault samples, inner ring fault samples,



(a) B1 (b) B2

(c) B3 (d) MB

FIGURE 5. Visualization of sample distribution in datasets B1, B2, B3, and MB using the UMAP tool.

and outer ring fault (OR) samples on the left side of Figure 5(c) explain the misclassification of some ball fault samples as inner ring fault samples and outer ring fault samples.

The distribution of samples in Figures 5(b) and 5(d) exhibits considerable disorder, which explains the generally poor recognition performance of various fault diagnosis methods on the B2 and MB datasets. In summary, a significant reason for the inaccurate recognition of fault samples by models such as 2DUMAP-KNN lies in the stronger similarity of the misclassified samples to those in the predicted category compared to their actual category.

V. CONCLUSION

This paper introduces two 2D linear dimensionality reduction algorithms based on UMAP and applies them to the field of fault diagnosis. The core idea revolves around learning a mapping matrix that enables new unknown samples to be directly projected from the high-dimensional space to the low-dimensional space. The proposed algorithms offer the following advantages:

- They address the inherent Out-of-Sample Problem associated with the UMAP algorithm;
- Compared to PUMAP, the 2DUMAP and 2DSUMAP methods exhibit lower time complexity;
- 2DSUMAP effectively leverages the label information of samples, thereby enhancing the discriminative capability of the mapping matrix and significantly improving the performance of 2DUMAP.

Experimental results on multiple bearing and gear datasets demonstrate that 2DUMAP and 2DSUMAP can accurately identify vibration signals associated with different types of faults. When compared to existing machine learning methods, our proposed approaches exhibit higher classification accuracy.

However, the research presented in this paper also has some limitations that merit further exploration and investigation. Both 2DUMAP and 2DSUMAP optimize the mapping matrix through an iterative process, resulting in relatively high time complexity. Therefore, further reducing the computational complexity to enhance the efficiency and practicality of the algorithms is necessary. Additionally, the metric used in 2DSUMAP has room for improvement.

COMPETING INTERESTS

The authors declare that there is no conflict of interest.

DATA AVAILABILITY

Data will be made available on request.

REFERENCES

[1] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.

[2] Y. Wu, Z. Fu, and J. Fei, "Fault diagnosis for industrial robots based on a combined approach of manifold learning, treelet transform and naive Bayes," *Rev. Sci. Instrum.*, vol. 91, no. 1, Jan. 2020, Art. no. 015116.

[3] Z. Zhu, Y. Lei, G. Qi, Y. Chai, N. Mazur, Y. An, and X. Huang, "A review of the application of deep learning in intelligent fault diagnosis of rotating machinery," *Measurement*, vol. 206, Jan. 2023, Art. no. 112346.

[4] Z. Mian, X. Deng, X. Dong, Y. Tian, T. Cao, K. Chen, and T. A. Jaber, "A literature review of fault diagnosis based on ensemble learning," *Eng. Appl. Artif. Intell.*, vol. 127, Jan. 2024, Art. no. 107357.

[5] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[7] S. Fan, X. Zhang, and Z. Song, "Imbalanced sample selection with deep reinforcement learning for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2518–2527, Apr. 2022.

[8] S. Ding, P. Zhang, E. Ding, A. Naik, P. Deng, and W. Gui, "On the application of PCA technique to fault diagnosis," *Tsinghua Sci. Technol.*, vol. 15, no. 2, pp. 138–144, Apr. 2010.

[9] S. Gajjar, M. Kulahci, and A. Palazoglu, "Real-time fault detection and diagnosis using sparse principal component analysis," *J. Process Control*, vol. 67, pp. 112–128, Jul. 2018.

[10] A. E. Chaleshtori and A. Aghaie, "A novel bearing fault diagnosis approach using the Gaussian mixture model and the weighted principal component analysis," *Rel. Eng. Syst. Saf.*, vol. 242, Feb. 2024, Art. no. 109720.

[11] J. Li, Z. Meng, N. Yin, Z. Pan, L. Cao, and F. Fan, "Multi-source feature extraction of rolling bearing compression measurement signal based on independent component analysis," *Measurement*, vol. 172, Feb. 2021, Art. no. 108908.

[12] S. Wang, Y. Wang, J. Tong, and Y. Chang, "Fault monitoring based on the VLSW-MADF test and DLPPCA for multimodal processes," *Sensors*, vol. 23, no. 2, p. 987, Jan. 2023.

[13] H. Chen, J. Wu, B. Jiang, and W. Chen, "A modified neighborhood preserving embedding-based incipient fault detection with applications to small-scale cyber–physical systems," *ISA Trans.*, vol. 104, pp. 175–183, Sep. 2020.

[14] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.

[15] M. Joswiak, Y. Peng, I. Castillo, and L. H. Chiang, "Dimensionality reduction for visualizing industrial chemical process data," *Control Eng. Pract.*, vol. 93, Dec. 2019, Art. no. 104189.

[16] J. An, P. Ai, C. Liu, S. Xu, and D. Liu, "Deep clustering bearing fault diagnosis method based on local manifold learning of an autoencoded embedding," *IEEE Access*, vol. 9, pp. 30154–30168, 2021.

[17] B. Zhang and P. Shang, "Distance correlation entropy and ordinal distance complexity measure: Efficient tools for complex systems," *Nonlinear Dyn.*, vol. 112, no. 2, pp. 1153–1172, Jan. 2024.

[18] H. Strange and R. Zwiggelaar, "A generalised solution to the out-of-sample extension problem in manifold learning," in *Proc. AAAI Conf. Artif. Intell.*, Aug. 2011, vol. 25, no. 1, pp. 471–476.

[19] A. W. Long and A. L. Ferguson, "Landmark diffusion maps (L-dMaps): Accelerated manifold learning out-of-sample extension," *Appl. Comput. Harmon. Anal.*, vol. 47, no. 1, pp. 190–211, Jul. 2019.

[20] G. Taskin and M. M. Crawford, "An out-of-sample extension to manifold learning via meta-modeling," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5227–5237, Oct. 2019.

[21] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, Dec. 2003, pp. 153–160.

[22] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Jan. 2005, pp. 1208–1213.

[23] K. Bunte, M. Biehl, and B. Hammer, "A general framework for dimensionality-reducing data visualization mapping," *Neural Comput.*, vol. 24, no. 3, pp. 771–804, Mar. 2012.

[24] H. Qiao, P. Zhang, D. Wang, and B. Zhang, "An explicit nonlinear mapping for manifold learning," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 51–63, Feb. 2013.

[25] B. Liu, S.-X. Xia, F.-R. Meng, and Y. Zhou, "Extreme spectral regression for efficient regularized subspace learning," *Neurocomputing*, vol. 149, pp. 171–179, Feb. 2015.

[26] A. Gisbrecht, A. Schulz, and B. Hammer, "Parametric nonlinear dimensionality reduction using kernel t-SNE," *Neurocomputing*, vol. 147, pp. 71–82, Jan. 2015.

[27] M. Tai, M. Kudo, A. Tanaka, H. Imai, and K. Kimura, "Kernelized supervised Laplacian eigenmap for visualization and classification of multi-label data," *Pattern Recognit.*, vol. 123, Mar. 2022, Art. no. 108399.

[28] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[29] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 17, Dec. 2004, pp. 1569–1576.

[30] L. Zhu and S.-A. Zhu, "Face recognition based on two dimensional locality preserving projections," *J. Image Graph.*, vol. 12, no. 11, pp. 2043–2047, 2007.

[31] W. Li, M. Qiu, Z. Zhu, B. Wu, and G. Zhou, "Bearing fault diagnosis based on spectrum images of vibration signals," *Meas. Sci. Technol.*, vol. 27, no. 3, Mar. 2016, Art. no. 035005.

[32] J. Yu and J. Liu, "Two-dimensional principal component analysis-based convolutional autoencoder for wafer map defect detection," *IEEE Trans. Ind. Electron.*, vol. 68, no. 9, pp. 8789–8797, Sep. 2021.

[33] Z. Luo, Z. Su, F. Tan, H. Ruixing, X. Hu, and B. Xing, "Machinery fault diagnosis based on weighted 2D fault feature extraction and multi-level information fusion," in *Proc. Int. Conf. Sens., Diag., Prognostics, Control (SDPC)*, Aug. 2020, pp. 296–302.

[34] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell, "Dimensionality reduction for visualizing single-cell data using UMAP," *Nature Biotechnol.*, vol. 37, no. 1, pp. 38–44, Jan. 2019.

[35] M. Vermeulen, K. Smith, K. Eremin, G. Rayner, and M. Walton, "Application of uniform manifold approximation and projection (UMAP) in spectral imaging of artworks," *Spectrochimica Acta A, Mol. Biomolecular Spectrosc.*, vol. 252, May 2021, Art. no. 119547.

[36] D. Milošević, A. S. Medeiros, M. S. Piperac, D. Cvijanović, J. Soininen, A. Milosavljević, and B. Predić, "The application of uniform manifold approximation and projection (UMAP) for unconstrained ordination and classification of biological indicators in aquatic ecology," *Sci. Total Environ.*, vol. 815, Apr. 2022, Art. no. 152365.

[37] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[38] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.

[39] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[40] H. Huang and N. Baddour, "Bearing vibration data collected under time-varying rotational speed conditions," *Data Brief*, vol. 21, pp. 1745–1749, Dec. 2018.

[41] A. P. Daga, A. Fasana, S. Marchesiello, and L. Garibaldi, "The politecnico di Turin rolling bearing test rig: Description and analysis of open access data," *Mech. Syst. Signal Process.*, vol. 120, pp. 252–273, Apr. 2019.

[42] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.

[43] P. Cao, S. Zhang, and J. Tang, "Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning," *IEEE Access*, vol. 6, pp. 26241–26253, 2018.

[44] A. C. Neto and A. L. M. Levada, "Probabilistic nearest neighbors based locality preserving projections for unsupervised metric learning," *J. Universal Comput. Sci.*, vol. 30, no. 5, p. 603, 2024.

[45] T. Sainburg, L. McInnes, and T. Q. Gentner, "Parametric UMAP embeddings for representation and semisupervised learning," *Neural Comput.*, vol. 33, no. 11, pp. 2881–2907, Aug. 2021.

[46] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[47] Y. Zhu, G. Li, R. Wang, S. Tang, H. Su, and K. Cao, "Intelligent fault diagnosis of hydraulic piston pump combining improved LeNet-5 and PSO hyperparameter optimization," *Appl. Acoust.*, vol. 183, Dec. 2021, Art. no. 108336.

[48] T. Jin, C. Yan, C. Chen, Z. Yang, H. Tian, and S. Wang, "Light neural network with fewer parameters based on CNN for fault diagnosis of rotating machinery," *Measurement*, vol. 181, Aug. 2021, Art. no. 109639.

[49] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.

[50] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.

[51] Y.-L. He, Y. Zhao, X. Hu, X.-N. Yan, Q.-X. Zhu, and Y. Xu, "Fault diagnosis using novel AdaBoost based discriminant locality preserving projection with resamples," *Eng. Appl. Artif. Intell.*, vol. 91, May 2020, Art. no. 103631.

BENCHAO LI is currently pursuing the master's degree with the College of Computer and Information Science, Chongqing Normal University, Chongqing, China. He is doing machine learning, dimensionality reduction, and computer vision research.

YUANYUAN ZHENG is currently pursuing the master's degree with the College of Computer and Information Science, Chongqing Normal University, Chongqing, China. She is doing machine learning, dimensionality reduction, and computer vision research.

RUISHENG RAN (Member, IEEE) received the B.S. degree in mathematics from Chongqing Normal University, Chongqing, China, and the M.S. degree in computational mathematics and the Ph.D. degree in computer application technology from the University of Electronic Science and Technology of China, Chengdu, China. He is currently a Professor with the College of Computer and Information Science, Chongqing Normal University. He is doing pattern recognition and computer vision research.