2024 6th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI)

Discriminant Neighborhood Preserving Embedding with Autoencoder for fault diagnosis

1st Ting Wang*
College of Computer and Information Science
Chongqing Normal University
Chongqing, China
2443610915@qq.com*

2nd Yisha Xie
College of Computer and Information Science
Chongqing Normal University
Chongqing, China
1698419083@qq.com

2024 6th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI) | 979-8-3315-4169-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/RICAI64321.2024.10911575

***Abstract*—With the development of modern industry, fault diagnosis has become a hot research area. The Discriminant Neighborhood Preserving Embedding (DNPE) algorithm has demonstrated excellent performance in handling complex industrial data. However, DNPE only preserves the local structure and discriminative information after dimensionality reduction, without the ability to recover the low-dimensional embedded data back to its original form. This lack of recovery capability may lead to the loss of important information in the data. Inspired by the reconstruction mechanism of autoencoders, a new NPE algorithm, DNPE-AE (Discriminant Neighborhood Preserving Embedding with Autoencoder), is proposed. Experiments on the bearing fault dataset provided by Case Western Reserve University (CWRU), show that DNPE-AE outperforms other methods in fault recognition and classification.**

***Index Terms*—discriminant neighborhood preserving embedding, autoencoder, manifold learning, fault diagnosis**

I. Introduction

With the development of modern industry and the widespread use of sensors, industrial fault data has become increasingly complex[1]. When a fault occurs in the production process, it can pose a serious threat to people's lives and property safety. Efficient fault diagnosis technology not only helps ensure safety but also facilitates timely maintenance of vulnerable components, such as bearings[2]. Data-driven fault diagnosis methods have become a research hotspot in the field of fault diagnosis due to their ability to perform efficient data extraction relying solely on data and models[3].

The primary challenge faced by data-driven fault diagnosis methods is how to extract more meaningful information from vast amounts of process data characterized by high dimensionality, nonlinearity, and strong interdependencies[4]. Fortunately, manifold learning (ML) methods have shown excellent performance in handling such high-dimensional data. Therefore, the application of manifold learning (ML) methods for feature extraction has significantly increased in the field of fault diagnosis[5]. Some classic manifold learning methods are frequently applied in the field of fault diagnosis[6]. For example, Q. Jiang et al. proposed a novel Laplacian Eigenmaps (LE) method for mechanical fault diagnosis, which improves the denoising performance of the LE algorithm by reconstructing the neighborhood relationships between sample data[7]. H. Zhang et al. proposed a fault detection method based on bi-kernel t-SNE (t-distributed Stochastic Neighbor Embedding) for semiconductor etching processes[8]. Qunxiong Zhu et al. introduced a fault diagnosis method based on Discriminant Neighborhood Preserving Embedding (DNPE), which replaces the distance metric with the Mahalanobis distance, addressing the issues of classification accuracy and data overlap[9].

The fault diagnosis methods based on manifold learning mentioned above are capable of preserving local neighborhood information in some cases, and in others, both local neighborhood and discriminative information. However, none of these methods consider whether the low-dimensional embedded data can be recovered back to its original form after dimensionality reduction. This lack of recovery capability can lead to the loss of important information from the original data, ultimately reducing classification performance.

In this paper, a new DNPE algorithm, Discriminant Neighborhood Preserving Embedding with Autoencoder (DNPE-AE), is proposed. DNPE-AE addresses the issues of the original DNPE algorithm, which lacks the ability to restore the embedded data to its original space. The DNPE-AE algorithm utilizes the reconstruction mechanism of the Autoencoder to enable the embedded data to be restored to its original form, thereby enhancing the fault diagnosis performance of the original NPE method. The experimental results on the CWRU bearing fault dataset indicate that the proposed DNPE-AE-based fault diagnosis method exhibits excellent fault classification performance and strong robustness for complex data.

II. Proposed Methodology

A. Framework and the objective function

In this section, a new framework for the DNPE method is proposed. Based on a linear autoencoder, DNPE-AE consists of two stages: The first stage involves feature extraction using the original DNPE algorithm, which reduces the dimensionality of the data while preserving the local neighborhood information and discriminative features of the original samples. The second stage employs the reconstruction mechanism of the autoencoder to enable the embedded data to be restored to its original form, recovering the data back to the original high-dimensional space. The framework of DNPE-AE is shown in Fig.1.

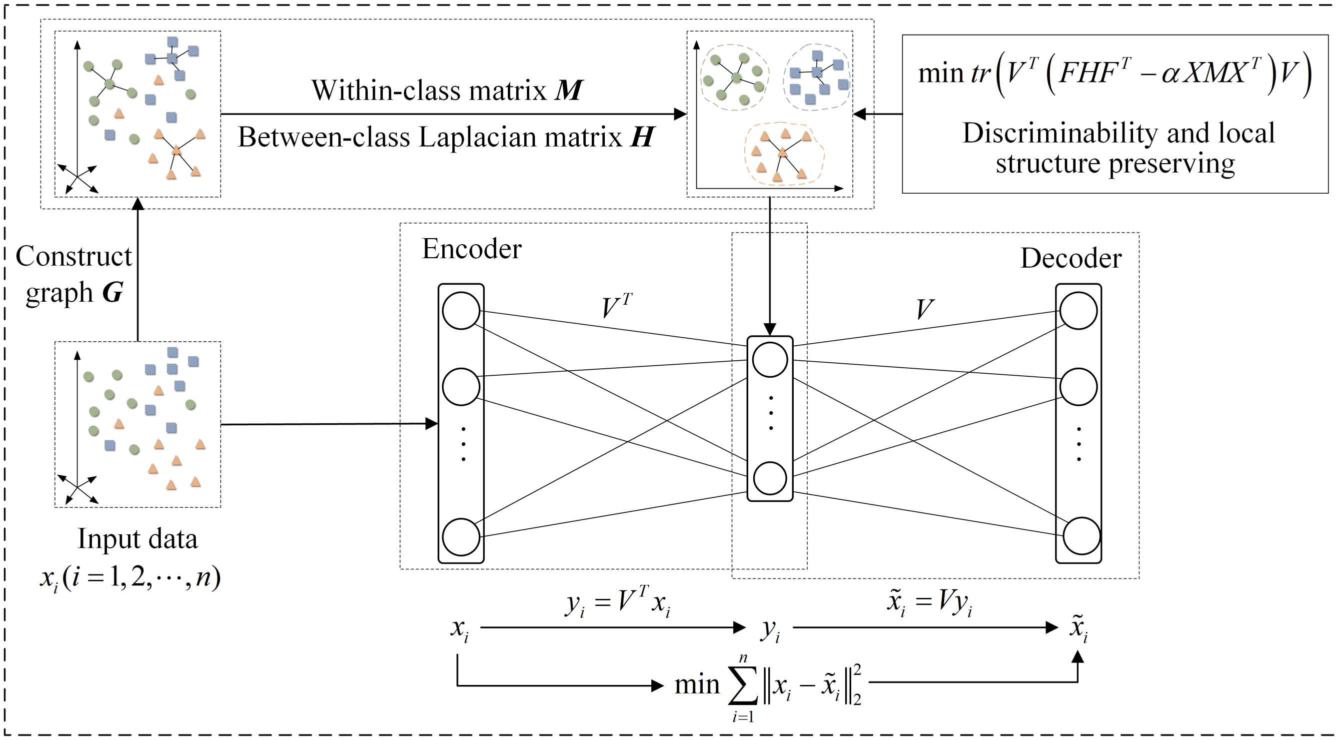

Fig. 1. The framework of the proposed DNPE-AE method.

Assume the original high-dimensional data is represented as $\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_N]$ in the space $\mathbb{R}^D$, each $\boldsymbol{x}_i$ belongs to a sample class $\{\boldsymbol{X}_1,\boldsymbol{X}_2,...,\boldsymbol{X}_C\}$. Let $\boldsymbol{F}=\left[\boldsymbol{f}_1,\boldsymbol{f}_2,...,\boldsymbol{f}_C\right]$, where $\boldsymbol{f}_i$ represents the center of the i-th class. First, the nearest neighbor graph is constructed using the k-nearest neighbor method. The first stage of DNPE-AE is the traditional DNPE projection, where the embedding data $\boldsymbol{Y}=[\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_N]$ is obtained through a linear mapping $\boldsymbol{y}_i=\boldsymbol{V}^T\boldsymbol{x}_i$. The mapping maximizes the between-class distance while minimizing the within-class distance. Therefore, the first term of the DNPE-AE objective function is achieved through the objective function of DNPE, as follows:

$$\mathcal{L}_1=-tr\left(\boldsymbol{V}^T\left(\boldsymbol{F}\boldsymbol{H}\boldsymbol{F}^T-\alpha\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T\right)\boldsymbol{V}\right) \tag{1}$$

Where α is the balance coefficient. The Laplacian matrix $\boldsymbol{H}$ is defined as $\boldsymbol{H}=\boldsymbol{E}-\boldsymbol{B}$, where $\boldsymbol{B}$ is the inter-class weight matrix and is given by $\boldsymbol{B}_{ij}=\exp\left(-\left\|\boldsymbol{f}_i-\boldsymbol{f}_j\right\|^2/t\right)$, $\boldsymbol{t}$ is an empirical parameter. Additionally, the matrix elements $\boldsymbol{E}_{ii}$ are defined as $\boldsymbol{E}_{ii}=\sum_i\boldsymbol{B}_{ij}$. $\boldsymbol{M}=(\boldsymbol{I}-\boldsymbol{W})^T(\boldsymbol{I}-\boldsymbol{W})$, Where $\boldsymbol{W}$ is the intra-class weight matrix, and $\boldsymbol{W}_{ij}^C=\exp\left(-\left\|\boldsymbol{x}_i^C-\boldsymbol{x}_j^C\right\|^2/t\right)$.

The second stage of DNPE-AE is the reconstruction stage of the embedded data, where the low-dimensional embedded data is restored to its original high-dimensional form using the decoding mechanism of a linear autoencoder. Let $\tilde{\boldsymbol{x}}_i$ represent the reconstruction of the data point $\boldsymbol{x}_i$, and $\boldsymbol{V}^*$ denote the weight matrix of the decoder. The reconstruction process can be formulated as follows:

$$\tilde{\boldsymbol{x}}_i=\boldsymbol{V}^*\boldsymbol{y}_i \tag{2}$$

Based on the simplified model in reference [10], the weight matrix $\boldsymbol{V}^*$ is rewritten as $\boldsymbol{V}^*=\left(\boldsymbol{V}^T\right)^T=\boldsymbol{V}$. Therefore, formula (2) can be rewritten as:

$$\tilde{\boldsymbol{x}}_i=\boldsymbol{V}\boldsymbol{y}_i=\boldsymbol{V}\boldsymbol{V}^T\boldsymbol{x}_i \tag{3}$$

The objective of the second stage of DNPE-AE is to achieve "optimal reconstruction" by minimizing the reconstruction error between the reconstructed data $\tilde{\boldsymbol{x}}_i$ and the original data $\boldsymbol{x}_i$. This objective ensures that the low-dimensional embedded data is restored to the original high-dimensional space while preserving the important information of the data. Therefore, the second term of the DNPE-AE objective function is to minimize the reconstruction error, which can be expressed as:

$$\begin{aligned}\mathcal{L}_2&=\sum_{i=1}^{n}\|\boldsymbol{x}_i-\tilde{\boldsymbol{x}}_i\|_2^2\\&=\sum_{i=1}^{n}\left\|\boldsymbol{x}_i-\boldsymbol{V}\boldsymbol{V}^T\boldsymbol{x}_i\right\|_2^2\\&=\sum_{i=1}^{n}\left\|\left(\boldsymbol{I}-\boldsymbol{V}\boldsymbol{V}^T\right)\boldsymbol{x}_i\right\|_2^2\\&=tr\left(\left(\boldsymbol{I}-\boldsymbol{V}\boldsymbol{V}^T\right)\boldsymbol{X}\boldsymbol{X}^T\left(\boldsymbol{I}-\boldsymbol{V}\boldsymbol{V}^T\right)^T\right)\end{aligned} \tag{4}$$

The goal of DNPE-AE is to find the optimal projection matrix $\boldsymbol{V}$ such that the low-dimensional embedded data $\boldsymbol{Y}$, obtained after projection, not only preserves the local neighborhood structure and discriminative information but also has the ability to recover to the original high-dimensional form. Therefore, combining equation (1) and equation (4), the objective function of DNPE-AE can be formulated as the minimization of the following:

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_1 + \gamma\mathcal{L}_2 \\ &= -tr\left(\boldsymbol{V}^T\left(\boldsymbol{F}\boldsymbol{H}\boldsymbol{F}^T - \alpha\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T\right)\boldsymbol{V}\right) \\ &+ \gamma tr\left(\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^T\right)\boldsymbol{X}\boldsymbol{X}^T\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^T\right)^T\right)\end{aligned} \tag{5}$$

Where γ is a parameter that balances the two terms, controlling the trade-off between maintaining the local structure and ensuring the data's recovery to its original form. The optimization process of parameters α and γ is detailed in the next subsection.

B. Optimization

Since the objective function (equation (5)) of the proposed DNPE-AE method is nonlinear, the optimal projection matrix $\boldsymbol{V}$ cannot be directly obtained using eigenvalue decomposition. Instead, we employ gradient descent to find the optimal projection matrix. As there are multiple optimization targets (the projection matrix $\boldsymbol{V}$, parameter α and γ), the multivariate gradient descent method is used to simultaneously optimize the projection matrix $\boldsymbol{V}$, parameter α and γ.

First, the gradient of the DNPE-AE objective function is computed. The formula for the gradient is as follows:

$$\begin{aligned}&\frac{\partial\mathcal{L}}{\partial\boldsymbol{V}} = \frac{d\mathcal{L}_1}{d\boldsymbol{V}} + \gamma\frac{d\mathcal{L}_2}{d\boldsymbol{V}}, \\ &\frac{\partial\mathcal{L}}{\partial\alpha} = tr\left(\boldsymbol{V}^T\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T\boldsymbol{V}\right), \\ &\frac{\partial\mathcal{L}}{\partial\gamma} = tr\left(\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^T\right)\boldsymbol{X}\boldsymbol{X}^T\left(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^T\right)^T\right).\end{aligned} \tag{6}$$

Where,

$$\begin{aligned}&\frac{d\mathcal{L}_1}{d\boldsymbol{V}} = 2\left(\alpha\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^T - \boldsymbol{F}\boldsymbol{H}\boldsymbol{F}^T\right)\boldsymbol{V}, \\ &\frac{d\mathcal{L}_2}{d\boldsymbol{V}} = -4\big(\boldsymbol{I} - \boldsymbol{V}\boldsymbol{V}^T\big)\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{V}.\end{aligned} \tag{7}$$

Then, using the computed gradient (6), the following formulas are used to update the matrix $\boldsymbol{V}$, parameter α and γ until the optimal matrix $\boldsymbol{V}$, parameter α and γ) are found:

$$\begin{aligned}\boldsymbol{V}_{t+1} &= \boldsymbol{V}_t - \beta\frac{\partial\mathcal{L}}{\partial\boldsymbol{V}_t} \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{\alpha}_t - \beta\frac{\partial\mathcal{L}}{\partial\boldsymbol{\alpha}_t} \\ \boldsymbol{\gamma}_{t+1} &= \boldsymbol{\gamma}_t - \beta\frac{\partial\mathcal{L}}{\partial\boldsymbol{\gamma}_t}\end{aligned} \tag{8}$$

The parameter β represents the step size in the gradient direction, and t denotes the number of updates.

Through the above optimization process, we obtain the optimal projection matrix $\boldsymbol{V}$, as well as the optimal parameters α and γ. After the linear mapping $\boldsymbol{y}_i = \boldsymbol{V}^T\boldsymbol{x}_i$, we obtain the optimal low-dimensional embedded data $\boldsymbol{Y}$. The features $\boldsymbol{Y}$ extracted using the DNPE-AE method not only preserve the local neighborhood structure and discriminative information of the original data $\boldsymbol{X}$, but also possess the ability to be recovered back to the original high-dimensional space, thereby providing a better "representation" of the original data.

III. Experiments

In this section, to validate the effectiveness of the proposed DNPE-AE-based fault diagnosis method, we conducted experiments using the bearing dataset provided by Case Western Reserve University (CWRU). Additionally, we reproduced classic related methods such as Locally Preserving Projection (LPP) [11], Principal Component Analysis (PCA) [12], Discriminative Locally Preserving Projection (DLPP) [13], Maximum Margin Criterionand (MMC) [14], and DNPE for performance comparison. The performance metrics compared include fault classification accuracy, precision, recall, and F1-score. The dimensionality of the data after reduction was set to 13.

A. CWRU dataset

The bearing dataset provided by Case Western Reserve University (CWRU) was collected using a test bench consisting of a 1.5 kW motor, drive-end bearings, fan-end bearings, a torque sensor, dynamometer, accelerometers, and an electronic controller. The raw data includes four fault conditions: normal, inner race fault, outer race fault, and ball fault, with fault diameters of 7 mils, 14 mils, and 21 mils. In this experiment, the data was re-sampled to a size of 1 × 128. We selected five fault types from the original 10 faults to validate the algorithm performance, specifically fault 3, fault 5, fault 6, fault 7, and fault 8. For each fault type, there are 312 training samples and 624 testing samples.

B. Experimental Results and Analysis

First, Table I presents the classification accuracy, precision, recall, and F1 score of each algorithm on the CWRU dataset. From Table I, it can be seen that the proposed DNPE-AE method excels in all evaluation metrics, achieving 100% accuracy, precision, recall, and F1 score, significantly outperforming other methods. This demonstrates its strong capabilities in feature extraction and fault classification. In contrast, MMC and LPP show weaker performance, with accuracy and F1 score significantly lower than other methods. Although DLPP and DNPE perform well in terms of accuracy, with 97.77% and 95.25%, respectively, other classification metrics are slightly lower, highlighting their limitations in handling complex industrial data. Overall, the excellent performance of DNPE-AE suggests that it better adapts to industrial fault diagnosis tasks, with stronger generalization ability and robustness compared to other methods.

To further validate the performance of the algorithm, Fig.2 shows the colormaps for each algorithm. As seen from Fig.2, corresponding to the results in Table I, DNPE-AE achieves 100% classification accuracy. Other methods, however, exhibit

TABLE I
COMPARISON OF CLASSIFICATION PERFORMANCE(%) OF DIFFERENT METHODS ON DATASET CWRU

Method	**Accuracy**	**Precision**	**Recall**	**F1**
LPP	72.31	69.11	68.51	68.65
PCA	95.25	93.71	95.73	94.63
DLPP	97.77	97.88	97.26	97.56
MMC	34.78	43.83	35.60	38.39
DNPE	90.38	88.34	87.34	86.70
DLPP-AE	**100**	**100**	**100**	**100**

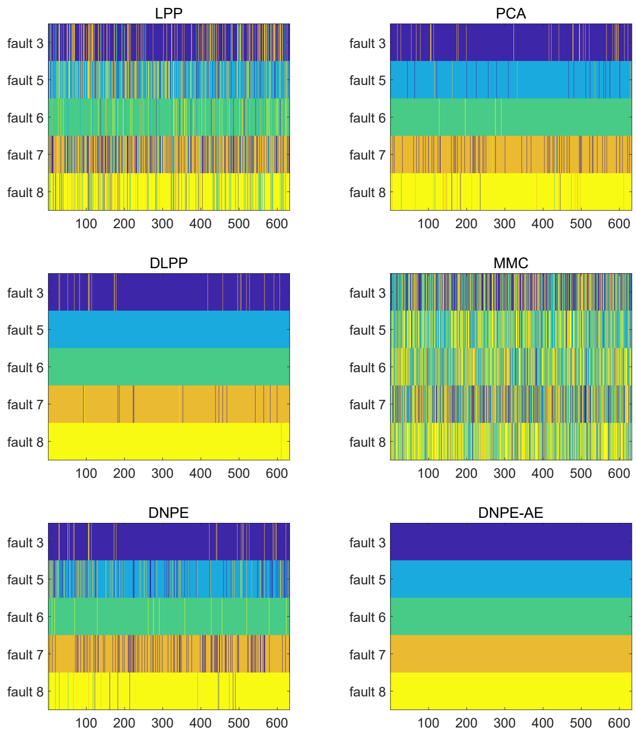

Fig. 2. The colormap on dataset CWRU.

some misclassifications. Although the purity of the colormap for DLPP is higher than that of other comparison methods, it is still significantly inferior to the proposed DNPE-AE method. Specifically, fault7 is often misclassified as fault3, while DNPE-AE can perfectly distinguish between these two faults. This is clearly demonstrated in the colormap, where a distinct separation between the two faults is observed, highlighting the superior fault diagnosis performance of DNPE-AE. This result underscores its strong classification capability and stability, especially when handling complex and similar fault types.

In summary, DNPE-AE outperforms other algorithms on the CWRU dataset, achieving 100% accuracy, precision, recall, and F1 score, significantly surpassing the other methods. In contrast, MMC and LPP show weaker performance, while DLPP, PCA and DNPE achieve good accuracy but slightly lower scores in other metrics. As shown in Fig.2, DNPE-AE can accurately distinguish between complex fault types, such as fault7 and fault3, which are often misclassified by other methods. Overall, DNPE-AE demonstrates superior generalization and stability in industrial fault diagnosis.

IV. CONCLUSION

This paper presents a novel Discriminant Neighborhood Preserving Embedding algorithm based on Autoencoder (DNPE-AE), which addresses the issue of the original DNPE algorithm's inability to recover the embedded data to its original form by utilizing the reconstruction mechanism of the autoencoder. This improvement allows the dimensionality-reduced embedded data to better retain the useful information from the original data, thereby enhancing fault diagnosis performance. Experimental results on the CWRU bearing dataset demonstrate that DNPE-AE outperforms other methods in fault diagnosis. However, this study also highlights areas for further research, such as the current method's applicability being limited to mechanical fault diagnosis. Future work will focus on extending this fault diagnosis method to chemical process data and exploring the application of more advanced autoencoders.

REFERENCES

[1] Zhiqin Zhu et al. "A review of the application of deep learning in intelligent fault diagnosis of rotating machinery". In: *Measurement* 206 (2023), p. 112346.

[2] Zheng Yang et al. "Autoencoder-based representation learning and its application in intelligent fault diagnosis: A review". In: *Measurement* 189 (2022), p. 110460.

[3] Huimin Zhao et al. "Feature extraction for data-driven remaining useful life prediction of rolling bearings". In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021), pp. 1–10.

[4] Huimin Zhao et al. "Feature extraction for data-driven remaining useful life prediction of rolling bearings". In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021), pp. 1–10.

[5] Jian Zhang, Jun Yu, and Dacheng Tao. "Local deep-feature alignment for unsupervised dimension reduction". In: *IEEE transactions on image processing* 27.5 (2018), pp. 2420–2432.

[6] Arash Moradzadeh et al. "Locating inter-turn faults in transformer windings using isometric feature mapping of frequency response traces". In: *IEEE Transactions on Industrial Informatics* 17.10 (2020), pp. 6962–6970.

[7] Quansheng Jiang et al. "An improved Laplacian Eigenmaps method for machine nonlinear fault feature extraction". In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 232.21 (2018), pp. 3833–3842.

[8] Haili Zhang et al. "Data visualization and fault detection using Bi-Kernel t-distributed stochastic neighbor embedding in semiconductor etching processes". In: *IEEE Transactions on Semiconductor Manufacturing* 35.3 (2022), pp. 522–531.

[9] Qunxiong Zhu et al. "Novel mahalanobis distance based fault diagnosis using discrimination neighborhood preserving embedding for industrial process". In: *2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS)*. IEEE. 2021, pp. 18–22.

[10] Yurii Nesterov. "A method for unconstrained convex minimization problem with the rate of convergence O (1/k2)". In: *Dokl. Akad. Nauk. SSSR*. Vol. 269. 3. 1983, p. 543.

[11] Shu Wang et al. "Fault Monitoring Based on the VLSW-MADF Test and DLPPCA for Multimodal Processes". In: *Sensors* 23.2 (2023), p. 987.

[12] Yongming Han et al. "An intelligent moving window sparse principal component analysis-based case based reasoning for fault diagnosis: Case of the drilling process". In: *ISA transactions* 128 (2022), pp. 242–254.

[13] Yan-Lin He et al. "Fault diagnosis using novel AdaBoost based discriminant locality preserving projection with resamples". In: *Engineering Applications of Artificial Intelligence* 91 (2020), p. 103631.

[14] Xiaohang Jin et al. "Motor bearing fault diagnosis using trace ratio linear discriminant analysis". In: *IEEE Transactions on Industrial Electronics* 61.5 (2013), pp. 2441–2451.