---

**Article:**
Yuan, S.-Z. orcid.org/0009-0009-1885-4542, Liu, Z.-H. orcid.org/0000-0002-6597-4741, Wei, H.-L. orcid.org/0000-0002-4704-7346 et al. (3 more authors) (2023) A variational auto-encoder based multi-source deep domain adaptation model using optimal transport for cross-machine fault diagnosis of rotating machinery. IEEE Transactions on Instrumentation and Measurement. ISSN 0018-9456

https://doi.org/10.1109/tim.2023.3331436

---

Final Accepted Manuscript (19 Oct 2023) for publication by *IEEE Transactions on Instrumentation & Measurement*

# A Variational Auto-encoder based Multi-Source Deep Domain Adaptation Model Using Optimal Transport for Cross-Machine Fault Diagnosis of Rotating Machinery

Shi-Zheng Yuan, Zhao-Hua Liu, *Senior Member, IEEE*, Hua-Liang Wei, Lei Chen, Ming-Yang Lv, and Xiao-Hua Li

***Abstract*— In recent years, most existing domain-adapted bearing fault diagnoses for rotating machinery are designed to decrease domain drifts for various operating conditions with an assumption that sufficient tag data are available. To overcome data scarcity, a possible solution is to use fault information of other machines of the same category to diagnose the status of a target machine (i.e., cross-machine diagnosis). This paper proposes a variational auto-encoder based multi-source deep domain adaptation model using optimal transport for cross-machine fault diagnosis of rotating machinery (named MDVAEOT). This is fundamentally different from most diagnostic models where both train and test data belong to the same machine. Firstly, it uses unlabeled samples of the machines to be diagnosed to establish the target dataset and faulty samples of machines of the same category (containing labels) to form the source dataset. Additionally, the method performs feature extraction on the dataset using variational auto-encoder networks and improves the reliability of extracted data features by the approximation of fixed probability. Finally, to shrink cross-machine differences between the two domains, we introduce optimal transport (OT) theory. OT distance is used to shares fault-related features between the two domains mentioned above to complete the cross-machine diagnosis task. Better accuracy and timeliness are offered by this proposed means compared to other existing intelligent methods in this field.**

***Index Terms*— Fault diagnosis, rotating machinery, variational auto-encoder, domain adaptation, optimal transport, cross-machine.**

## I. Introduction

EFFECTIVE fault diagnosis applied to rotating machinery is a prominent way for risk reduction of serious damage, which is essential for improving engineering reliability and ensuring system safety [1], [2]. With the increasing demands of modern industry on system safety and economic costs, the development of effective and reliable fault diagnosis methods has gradually become one of the most important research topics [3].

Fault diagnosis methods are able to be broadly categorized as four sorts: 1) model-based methods, 2) signal analysis-based methods [4], 3) feature extraction and machine learning-based methods [5], and 4) transfer and deep learning-based methods [6]. Model-based methodologies usually start with modelling the input-output relationship, together with some important environmental conditions, relating to an object or process of interest. Although such methods are highly interpretable, they sometimes require a lot of expert knowledge of the diagnosed object and its environment. Methods based on signal analysis mainly utilize time domain, frequency domain and time-frequency domain analysis of the signals of interest to discern the state of the underlying system. Such methods can usually improve the accuracy to a certain extent, but they may cost a lot of labor to process data. Methods based on feature extraction and machine learning mainly use traditional signal processing methods to extract features and then use machine learning models (e.g., k-means, SVM) to discriminate the processed features. This type of approaches provides some good ideas of using machine learning models, but the performance of feature engineering and machine learning models extremely guides the precision of the diagnosis.

With the strong development of data-driven technologies, many intelligent diagnostic methodologies have been presented that rely on deep learning in such a diagnostic field. Such methods extract feature-hidden depth structures by multiple training [7]. For example, Peng et al. [8] improved a multi-scale multi-branch convolutional neural network to reduce the effect of noise on diagnosis. Tang et al. [9] used the complete ensemble empirical mode decomposition, fast Fourier transform (FFT), and the recursive feature elimination (RFE) to select the optimal feature subset. The feature subsets extracted by this method are all used to train deep networks, and the resulting network models achieve good accuracy. Hou et al. [10] used multipoint optimal minimum entropy inverse fold product and sparse operations to increase the signal-to-noise ratio for fault feature enhancement and robust diagnosis. Yang et al. [11] used conditional generative adversarial networks (CGAN) to generate new samples with similar data

Manuscript received September 22, 2023 and accepted October 19, 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61972443, in part by the National Key Research and Development Project of China under Grant 2019YFE0105300, in part by the Hunan Provincial Key Research and Development Project of China under Grant 2022WK2006.

S.-Z Yuan, Z.-H Liu, L. Chen, M.-Y Lv and X.-H Li are with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China (e-mail: ysz@mail.hnust.edu.cn; zhaohualiu2009@hotmail.com; chenlei@hnust.edu.cn; 1040133@hnust.edu.cn; lixiaohua_0227@163.com).

H. -L. Wei is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, U.K. (e-mail: w.hualiang@sheffield.ac.uk).

distribution. A two-dimensional convolutional neural network (2D-CNN) was also adopted to address the two-dimensional grey-scale images converted from one-dimensional data. This method improves the accuracy in small samples.

In addition, other methods such as domain adaptation [12] and transfer learning [13] have also been introduced to overcome the imbalance issue in data distributions between training and test data due to different machine conditions (e.g. different loads, noise). In domain adaptation and transfer learning, the training and test data sets are usually called the source and target domains, respectively. Domain adaptation and transfer learning can reduce domain drifts in data distributions of two domains and enable related knowledge to be shared across domains, thus improve the diagnostic performance [14]. Pan et al. [15] suggested a transfer component analysis-based (TCA) domain adaptation technique to shrink the feature differences between two domains. In [16], domain invariant features were extracted by joint distribution adaptation (JDA). In [17] an attention mechanism was incorporated into auto-encoders with different activation functions and used dynamic domain factors to assign weights to marginal and conditional distributions to enhance the diagnostic performance in various operating conditions. In [18] proposes a two-stage transfer adversarial network in combination with transfer learning to complete the diagnosis task of multiple new faults under the target domain. Hu et al. [19] used various statistic-based multi-scale sample entropies to improve the feature recognition of diverse failure modes under different operating conditions. In [20] minimizes the source and target domain feature differences by maximum mean difference. In [21] aligns the category distribution within each source-target domain pair and reduces the negative migration of the global domain alignment. This method achieves good results in the case of coexisting field and category inconsistencies. The transfer learning in the literature [22], [23] all introduced optimal transport theory to reduce domain drifts, demonstrating superiority in terms of accuracy of fault diagnosis.

The deep learning and transfer learning-based approaches mentioned above focus on cross-domain fault diagnosis for the same system (process, machine, etc.) under different operating conditions. Data used for training and testing models comes from the same system. It is usually assumed that the training dataset has sufficient labelled data. However, such an assumption may not always be reasonable for real diagnosis tasks in engineering. In real industry, available labelled data for machines, especially negative samples, is scarce or even non-existent. Recently, researchers have embarked on a new idea of using diagnostic expertise acquired from other relevant ones to discriminate the state of the targeted machine. In [24] uses an auto-coder network to project features from different devices into the same subspace and then uses maximum mean difference to minimize differences of the data distributions from different devices. In [25] proposes to extract features from different devices using the structure of convolutional networks and dynamically align the edge probability distribution and conditional probability distribution of the data by a weighting factor based on the maximum mean difference. In [26] processes the data through a shared feature extraction module and a private feature extraction module to ensure that the extracted features are representative and shared. Most of these methods are based on Maximum Mean Difference and some still need a low quantity of available labelled samples in the target domain. The literature review shows that cross-machine fault diagnosis is also becoming an emerging trend.

This paper proposes a variational auto-encoder based multi-source deep domain adaptation model using optimal transport for cross-machine fault diagnosis of rotating machinery (MDVAEOT). In the proposed method, a source dataset is built through usage of fault samples of the machine in the laboratory and a target dataset is built with unlabeled samples from another machine to be diagnosed in the same class. The shared fault-related features serve to diagnose the targeted device. It performs unsupervised learning with pseudo-labels without requiring the target domain samples to be labelled. Its major contributions are outlined below.

1) We propose a new variational auto-encoder based multi-source deep domain adaptation model using optimal transport for cross-machine fault diagnosis in rotating machinery. It can automate the extraction of cross-machine domain invariant features and perform class discriminant analysis to accomplish end-to-end fault diagnosis.
2) To the issue that labeled data for a single working condition may be insufficient, we fuse multi-source domain data in the source domain. Additionally, we use variational auto-encoder networks (VAE) to improve the reliability of extracted data features. This is, as far as we know, the very first use of this network in this field. Experiments show that the deep model of VAE can extract the category features more efficiently.
3) To solve the cross-machine diagnosis problem, we introduce optimal transport (OT) theory. OT distance is adopted to discriminate the discrepancies in the data distributions among source and target domains. This allows fault-related features to be efficiently shared to the target domain.

The rest of this article continues with the problem formulation in Section II. Section III describes the model structure. Section IV provides an experimental analysis of it. We conclude the paper in Section V.

## II. Problem Formulation

Rotating machinery, as the supporting mechanical rotating body of the equipment, is affected by many elements, such as variable operating conditions, different loads and changing noise. In practice, relatively small or scarce of the fault data is available under a single operating condition. Moreover, we often do not earn substantial available labeled data of the target machine, especially fault information. This makes it difficult to diagnose the status of the machine by making use of the tag data of the machine.

In addition, many new machines that are not in service before

do not have pre-existing status information. We still need to check their status for industrial safety after they come into use. However, the data distribution varies greatly from one machine to another, with e.g., different sampling frequencies, mechanical structures and resonance characteristics. These differences are much greater than the inter-domain differences of the data distribution on the same machine. Owing to the wider domain drift, the corresponding commonalities between the different machines become lesser, as shown in Fig. 1.

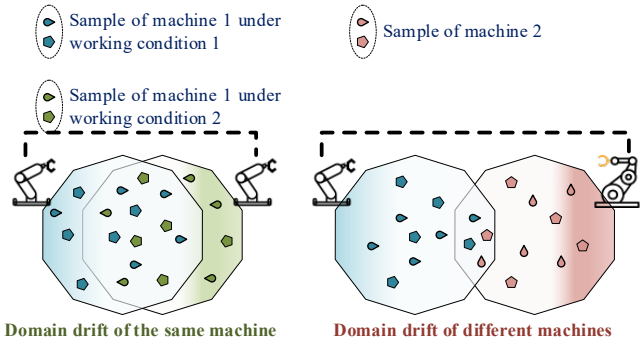



Fig. 1. The smaller domain drift in cross-machine fault diagnosis tasks.

Therefore, there is a need to design a new framework to create a diagnosis model that can be well utilized in the target machine by making use of the fault information of the machine in the same class. Moreover, it is expected that the framework can effectively utilize multi-condition data (i.e., multi-source domain data) for feature learning. This can compensate the shortcoming of scarce single-condition fault data and address the domain drift problem of fusing data of the same machine in multiple conditions.

The emphasis of the work in this article on the task of cross-machines diagnosis. The designed algorithm can learn transferable feature knowledge from the source machine and apply it to diagnosis of the target machine. In summary, this research is conducted with the following basic facts:

1) the source fault types contain the target fault types;
2) sufficient multi-source labeled samples are available for supervised learning;
3) sufficient unlabeled samples in the target domain are available for unsupervised learning;
4) the differences in the distribution of the same class of features across machines are smaller than the differences in the distribution of different classes of features on the same machine.

Fact 2 allows us to use labeled data from multiple conditions of the same machine to construct enough multi-source domain data for extracting valid and comprehensive fault category features to make up for the shortage of single condition data. This is more in line with industrial reality. Facts 1 and 4 ensure the validity of knowledge migration across machines under the condition of Fact 3.

## III. The Proposed Cross-Machine Fault Diagnosis Model Based on Variational Auto-Coder and Optimal Transport Distance

The flow chart of the suggested methodology is presented in Fig. 2. First, a feature extractor is constructed based on VAE [27] to extract features from the both domains. Then, the OT distance is used to shrink domain drifts across machines. Finally, a multi-category label discriminator is used to predict the labels.

### *A. Variational Auto-encoder-based Feature Extractor*

Good feature extraction is a prerequisite for accurate classification. The feature extractor constructed based on VAE is able to obtain effective features in multi-source domain scenarios with various working conditions of the same machine. VAE is a type of auto-coder network that combines neural networks and Bayesian formula with a basic encoder-decoder structure, as shown in Fig. 3. The structure of VAE is unique in that it has an additional latent layer. This latent layer fits the distribution of the input data to a Gaussian distribution, and then samples over this Gaussian distribution to reconstruct the original input data. Therefore, VAE can be used to estimate the probability distribution of the hidden variables.

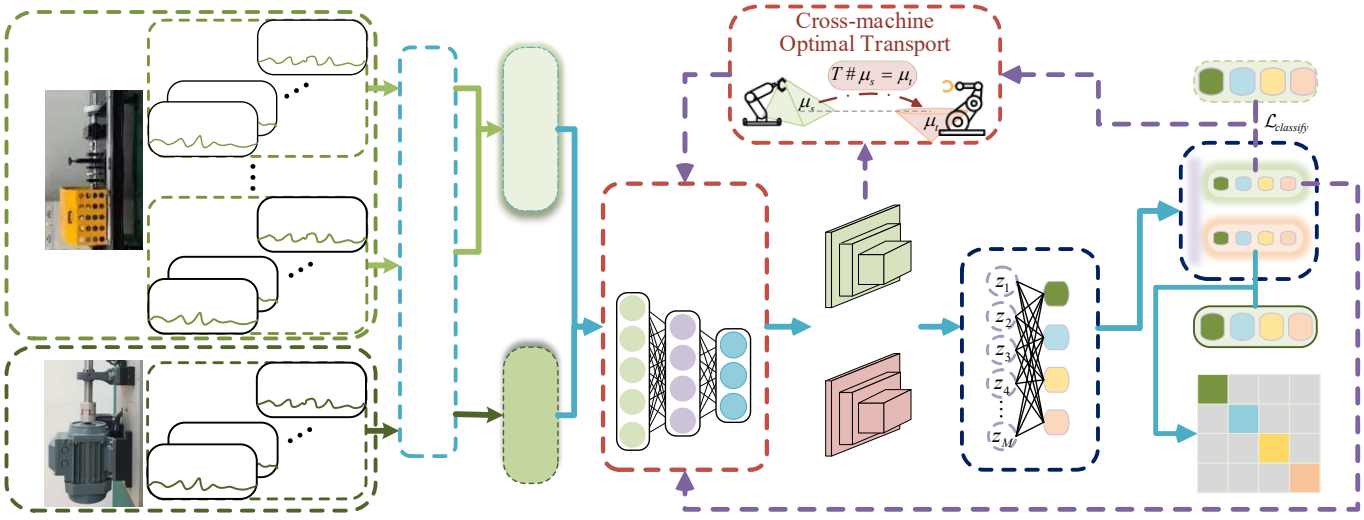



Fig. 2. Flow chart of the proposed cross-machine fault diagnosis model based on variational auto-coder and optimal transport distance.

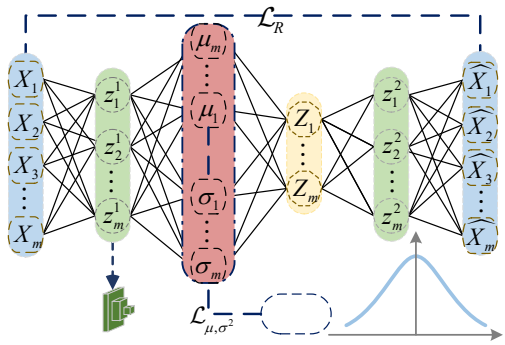


Fig. 3. Architecture of VAE-Based Feature Extractor.

Denote by $X=\{X_1, X_2, ..., X_m\}$ the whole sample set. For a sample $X_k$, assume that there exists a posterior distribution $P(Z|X_k)$ that is exclusive to $X_k$. Note that for a given sample $X_k$, it is assumed $P(Z|X_k)=N(0,I)$. From the Bayesian rule

$$P(Z) = P(Z|X_k)P(X_k) \tag{1}$$

It is reasonable to assume that a sampled $Z_k$ from the distribution $P(Z|X_k)$ can be reduced to $X_k$ by the assumption of exclusive membership. For the sample set $\{X_1, X_2, ..., X_m\}$, there should be $m$ independent, multivariate Gaussian distributions. To discover the variance and mean of the Gaussian distribution $P(Z|X_k)$ belonging exclusively to $X_k$, we construct two neural networks to compute as the following:

$$\mu_k = f_1(X_k), \qquad \log\sigma_k^2 = f_2(X_k) \tag{2}$$

The corresponding Gaussian distribution $P(Z|X_k)$ can then be determined using the results given in (2). In this way, we can sample a sample $Z_k$ from that distribution and reduce it to $X_k$ by defining a generator $\widehat{X_k} = g(Z_k)$ and a reconstruction loss function $D(\widehat{X_k}, X_k)^2$. To align $P(Z|X_k)$ to $N(0,I)$, we measure the KL distance of the Gaussian distribution from the standard normal distribution, i.e., $KL(N(\mu,\sigma^2)\|N(0,I))$, as follows:

$$\mathcal{L}_{\mu,\sigma^2} = \frac{1}{2}\sum_{i=1}^{d}(\mu_i^2 + \sigma_i^2 - \log\sigma_i^2 - 1) \tag{3}$$

where $d$ is the dimensionality of the sampled hidden variable $Z$. $\mu_i$ and $\sigma_i^2$ represent the $i$th component of the mean and variance vectors of Gaussian distributions.

Since the operation "sampling" cannot compute the gradient, and the result of sampling has a gradient, we introduce the reparameterization trick to sample a $Z$ from $N(\mu,\sigma^2)$. This trick is equivalent to sampling a $\varepsilon$ from $N(0,I)$, and making $Z = \mu + \varepsilon * \sigma$ .In this way, the "sampling" operation does not participate in gradient descent, but the sampled result does, making the whole model trainable.

To simplify the calculation, we perform dimensionality reduction on the original data. In the basic structure of VAE, we add 2 implicit layers $z^1$ and $z^2$ as shown in Figure 3. The mathematical relationships between the layers are as follows:

$$z^1 = f_{\varpi_1}(\omega_1 X + b_1), \quad \widehat{X} = f_{\varpi_2}(\omega_2 z^2 + b_2) \tag{4}$$

where $\omega$ and $b$ in $\varpi \triangleq \{\omega, b\}$ are the weight matrix and bias vector, and $f_\varpi$ is the activation function. Similar to the auto-coder network, the reconstructed loss function is as follows:

$$\mathcal{L}_R = \frac{1}{2m}\sum_{i=1}^{m}\left\|\widehat{X} - X\right\|_2 \tag{5}$$

Combining (5) with equation (3), the loss function of the feature extractor constructed based on the VAE is $\mathcal{L}_f = \mathcal{L}_{\mu,\sigma^2} + \mathcal{L}_R$.

### *B. Optimal Transport Distance for Cross-machine Adaptation*

In general, we minimize some predefined distance measures of domain differences to achieve knowledge sharing between domains. These distance measures include Kullback-Leibler divergence (KL) [28], maximum mean difference (MMD) [29], [30] and Jensen-Shannon divergence (JS) [31]. In the non-overlapping area of two distributions, KL may obtain meaningless results and JS can increase abruptly. MMD also requires cumbersome consideration of changes in the relevant weights of marginal and conditional probability distributions for significant increase of accuracy when measuring differences in data across machines. To effectively characterize the large differences in data between different machines of the same type, there is a need to develop more effective methods to weigh the discrepancies between the both domains. Optimal

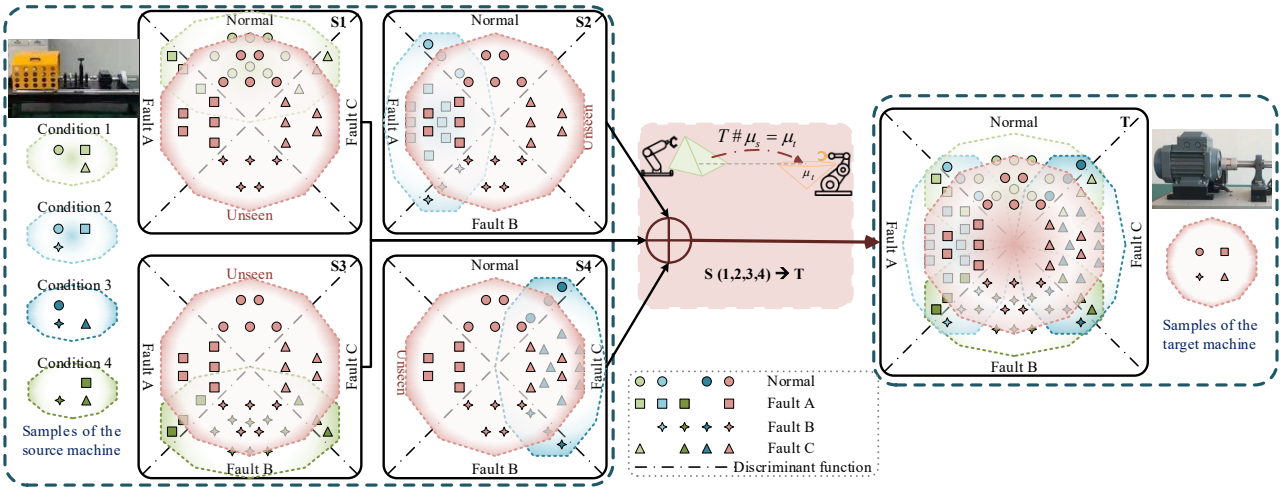



Fig. 4. Optimal transport distance for domain and category inconsistencies coexist in cross-machine fault diagnosis tasks.

transport (OT) theory is a robust probability measure [32] for quantifying and comparing two different probability distributions. The OT distance is calculated directly from the samples of the distribution, rather than performing density estimation. This distance is computed smoothly and does not have abrupt changes like JS or KL, which is well suited as a loss function for deep neural network models building. The OT distance uses linear programming, which incurs a higher computational cost. Fortunately, many effective algorithms [33] have been presented to enhance the simplicity of the computation. Considering gradient stability and ease of training, we choose the optimal transport to share knowledge across machines, as shown in Fig. 4.

Optimal transport distance can be defined as the distance between non-overlapping probability distributions. Let the loss function be c, consider the problem: how to minimize the cost of transporting a pile of sand of shape $\mu_s$ to a pit of shape $\mu_t$. This problem was first defined by the Monge problem:

$$\inf_T \left\{ \int c(x,T(x)) d\mu(x) \middle| T\#\mu_s = \mu_t \right\} \tag{6}$$

where $c(x,T(x))$ is the cost function. $T$ denotes a transport planning. $T\#\mu_s = \mu_t$ is the transport mapping from $\mu_s$ to $\mu_t$.

The Kantorovich problem defines a relaxation explanation of the optimal transport problem. There exists a transportation plan $\gamma \in \Gamma(\mu_s,\mu_t)$ with $\Gamma(\mu_s,\mu_t)$ being the set of all probabilistic couplings with marginal distributions $\mu_s$ and $\mu_t$. Then the p-order Wasserstein distance between $\mu_s$ and $\mu_t$ is defined as

$$W_p(\mu_s,\mu_t) := \inf_{\gamma\in\Gamma(\mu_s,\mu_t)} \left( \int c(x,y)^p d\gamma(x,y) \right)^{1/p} \tag{7}$$

The OT distance needs to be discretized to address cross-machine domain drifts. Let the marginal distributions of the samples in the both domains be $\mu_S$ and $\mu_T$, and the optimal transport plan $\gamma_*$ be given as follows:

$$\mu_S = \sum_{i=1}^{m} p_i^S \delta_{X_i^S}, \qquad \mu_T = \sum_{i=1}^{n} p_i^T \delta_{X_i^T} \tag{8}$$

$$\sum_{i=1}^{m} p_i^S = \sum_{i=1}^{n} p_i^T = 1 \tag{9}$$

$$\Gamma = \{ \gamma \in (\mathbb{R}^+)^{m\times n} \left| \gamma 1_n = \mu_S, \gamma^{\mathrm{T}} 1_m = \mu_T \right. \} \tag{10}$$

$$\gamma_* = \arg\min_{\gamma\in\Gamma} \langle c(X_i^S, X_i^T), \gamma \rangle_F \tag{11}$$

where $\delta_{X_i^S}$ and $\delta_{X_i^T}$ are the Dirac functions corresponding to $X_i^S$ and $X_i^T$, $\langle .,. \rangle_F$ is the Frobenius norm, $c(X_i^S, X_i^T)$ is the cost matrix. In this way, the source sample $X_i^S$ can be mapped into the target sample $\widehat{X_i^S}$ as the following:

$$\widehat{X_i^S} = \mathrm{T}_{\gamma_*}(X_i^S) = diag(\gamma_* 1_n)^{-1} \gamma_* X_t \tag{12}$$

The joint distribution optimal transport [34] is usually used when reducing the domain drift problem across machines with OT distance, considering both feature and label differences. The Euclidean distance is employed as a metric for the feature space. The cross-entropy loss metric is implemented in the label space. As a result, the loss function is defined as

$$c(X_i^S, Y_i^S, X_j^T, Y_j^T) = d_{Euc}(f(X_i^S), f(X_j^T)) + L(Y_i^S, Y_j^T) \tag{13}$$

where $d_{Euc}(.,.)$ refers to the Euclidean distance, $L(.,.)$ refers to the cross-entropy function, and $z = f(X)$ refers to the dimensionality reduction feature. Therefore, the optimization function equation (11) can be rewritten as:

$$\gamma_* = \arg\min_{\gamma_{i,j}\in\Gamma} \sum_{i,j} \gamma_{i,j} \begin{Bmatrix} d_{Euc}(f(X_i^S), f(X_j^T)) + \\ L(Y_i^S, \mathrm{g}(f(X_j^T))) \end{Bmatrix} \tag{14}$$

where $\mathrm{g}(f(X_j^T))$ represents the target pseudo labels.

### C. Multi-category Label Discriminator

The probability is derived from the Softmax function that a sample Xi belongs to a specific class as the following:

$$\begin{aligned} P(Y|X_i;\varpi) &= \left[ P(Y=1|X_i;\varpi), ..., P(Y=k|X_i;\varpi) \right]^{\mathrm{T}} \\ &= \frac{1}{\sum_{k=1}^{K} \exp(\varpi_k^{\mathrm{T}} X_i)} \begin{bmatrix} \exp(\varpi_1^{\mathrm{T}} X_i) \\ \vdots \\ \exp(\varpi_K^{\mathrm{T}} X_i) \end{bmatrix} \end{aligned} \tag{15}$$

For the multi-classification problem, we use the Softmax function as the label classifier. The cross-entropy is used as the loss function of the probability distribution of the predicted labels $y_{pre}$ and the probability distribution of the true labels $y_{auc}$ as the following:

$$\mathcal{L}_{classify} = \sum P(y_{pre}) \log\left(\frac{1}{P(y_{auc})}\right) \tag{16}$$

### D. Optimization Strategy

The parameters require to be optimized of the feature extractor *f*, label predictor *g* and OT plan $\gamma$, when the above method is adopted for cross-machine fault diagnosis. This is obtained by minimizing the objective function of Eq. (17) defined as

$$\begin{aligned} \min_{f,g,\gamma_{i,j}\in\Gamma} \mathcal{L}(f,g,\gamma_{i,j}) &= \sum_i \mathcal{L}_{classify}(Y_i^S, g(f(X_i^S))) + \\ \sum_{i,j} \gamma_{i,j} &\left\{ d_{Euc}(f(X_i^S), f(X_j^T)) + L(Y_i^S, \mathrm{g}(f(X_j^T))) \right\} \end{aligned} \tag{17}$$

Eq. (17) is able to be achieved by holding the value of $\gamma_{i,j}$ to solve for *f* and *g*, i.e., equation (18). Fixing the values of *f* and *g* to solve for $\gamma$, i.e., equation (19).

$$\min_{f,g,\gamma_{i,j}\in\Gamma} \mathcal{L}(f,g,\gamma_{i,j}) = \sum_i \mathcal{L}_{classify}(Y_i^S, g(f(X_i^S))) \tag{18}$$

$$\min_{f,g,\gamma_{i,j}\in\Gamma} \mathcal{L}(f,g,\gamma_{i,j}) = \sum_{i,j} \gamma_{i,j} \begin{Bmatrix} d_{Euc}(f(X_i^S), f(X_j^T)) + \\ L(Y_i^S, \mathrm{g}(f(X_j^T))) \end{Bmatrix} \tag{19}$$

In practice, it could be the case where not only the data distribution varies greatly between cross-machine tasks, but also the data magnitudes are not matched. Therefore, performing feature extraction and an OT scheme on source and target samples simultaneously inevitably suffers from data magnitude mismatch in the process of parameter optimization. We unify the data in the both domains to the same order of magnitude to reduce the convergence time. The gradient descent algorithm is adopted to optimally refresh the parameters of each part (Fig. 5). Conditions 1 and 2 merely

determine a certain number of iterations. Thus, the conditions are easily satisfied. Table I shows its specific parameters.

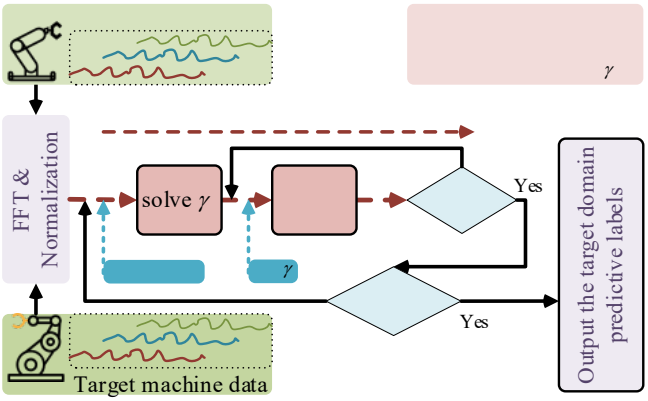



Fig. 5. Optimization strategy for the proposed cross-machine fault diagnosis model based on variational auto-coder and optimal transport distance.

TABLE I
PARAMETER SET OF PROPOSED METHOD

| Input layer | Input size | Connection method | Output size | Output layer |
|---|---|---|---|---|
| $X$ _layer | 600 | sigmoid | 400 | $z^1$_layer |
| $z^1$_layer | 400 | linear | 2 | $\mu,\sigma$ _layer |
| $\mu,\sigma$ _layer | 2 | $\mu+\varepsilon*\sigma$ | 2 | Z_layer |
| Z_layer | 2 | sigmoid | 400 | $z^2$_layer |
| $z^2$_layer | 400 | sigmoid | 600 | $\widehat{X}$ _layer |
| $z^1$_layer | 400 | linear | 4 | classifier_layer |
| Optimizer: | | Adam | | |

## IV. CASE VERIFICATION

### A. Descriptions of Datasets

***Dataset 1***: The CWRU [35] experiments were performed on the test stand on Fig. 6. Bearing failure types: inner ring failure (IF), rolling element failure (BF), and outer ring failure (OF). The failures were implanted with EDM on the motor bearings, with each type of failure measuring 0.007", 0.014", 0.021" and 0.028". The motor speed range was from 1720 to 1797 rpm. In this paper, the data from the accelerometer at the DE end was selected for the study, and the normal and fault data were sampled at 48 kHz and 12 kHz, with an acquisition time of 10 seconds. The normal data and fault data of 0.007" were selected for three operating states (i.e., 0hp, 1hp and 2hp) depending on the speed.

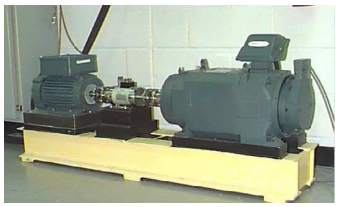


Fig. 6. Experimental equipment of CWRU.

***Dataset 2***: The PT100 test was performed on the test stand on Fig. 7. Selected bearing failure types: IF, BF and OF with 0.2mm cracks respectively. Motor speed range was 1800 to 2100 rpm, accelerometer sampling frequency was 48 kHz, and acquisition time was 2.78 seconds. Three operating states (i.e., 0hp, 1hp and 2hp) fault data were selected with normal data depending on the speed.

Fig. 7. Experimental equipment of PT100.

***Dataset 3***: The HD-FD-H-03X rotor gearbox comprehensive fault simulation test bench (Fig. 8) was used for the experiment. The selected bearing failure types: IF, BF and OF. Details are as follows: radial loading force of 1000 N, motor speed range of 1000 to 5700 rpm, sampling frequency of 16 kHz, sampling time of 4 seconds, and 3 sets of acceleration data (vertical, horizontal, and axial) for each working condition. Three operating conditions (i.e., 0 hp, 1 hp and 2 hp) fault data and normal data were selected according to the speed.

Fig. 8. Experimental equipment of HD-FD-H-03X.

**Task 1**: Take all samples of 3 working conditions in Dataset1 to form the multi-source dataset, and take the samples of 1 working condition in Dataset 2 to form the target dataset. Then reverse the process, and use the data of 3 conditions in Dataset 2 to form the multi-source dataset, and use the data of 1 condition in Dataset 1 to form the target dataset.

**Task 2**: Similar to **task 1**, only the Dataset 2 is replaced with the Dataset 3.

In this paper, every 1200 data points are considered as one sample. According to previous studies, the use of frequency domain signals is better than time series signals when using vibration data for analysis. Therefore, we use the fast Fourier transform (FFT) to process the vibration data in the frequency domain. The dimension 600 of the calculated one-sided frequency amplitude is employed as the final input dimension of the training model. The data set is disrupted before training.

Let $x_m^{S_n}$ and $y_m^{S_n}$ denote the $m$th sample with label in the $n$th working condition in the source domain, and $x_n^T$ denote the $m$th sample of the target domain data. The source domain dataset with its label set and the target domain dataset are

$$D_S = \{x_1^{S_1}, x_2^{S_1}, ..., x_m^{S_1}, x_1^{S_2}, x_2^{S_2}, ..., x_m^{S_2}, ..., x_1^{S_n}, x_2^{S_n}, ..., x_m^{S_n}\},$$

$$L_S = \{y_1^{S_1}, y_2^{S_1}, ..., y_m^{S_1}, y_1^{S_2}, y_2^{S_2}, ..., y_m^{S_2}, ..., y_1^{S_n}, y_2^{S_n}, ..., y_m^{S_n}\} \quad \text{and}$$

$$D_T = \{x_1^T, x_2^T, ....., x_n^T\}.$$

The diagnostic tasks are shown in the first and second columns of Table II, and the both domains of the diagnostic tasks contain a total of 4 status categories of normal and fault. Task 1_1 refers to the first subtask of Task 1: D1_0_1_2 hp ->D2_0 hp, i.e., use the samples of 3 working conditions of

Dataset1 to diagnose the samples of working condition 0 hp in Dataset 2.

TABLE II
ACCURACY (%) OF TASKS

| Diagnostic task | Domain shift | SVM | CORAL | TCA | JDA | SAEOT | **Proposed** |
|---|---|---|---|---|---|---|---|
| Task 1_1 | D1_0_1_2 hp ->D2_0 hp | 50.0 | 37.5 | 25.3 | 25.4 | 96.4 | **100** |
| Task 1_2 | D1_0_1_2 hp ->D2_1 hp | 50.0 | 37.5 | 25.3 | 25.4 | 96.5 | **98.9** |
| Task 1_3 | D1_0_1_2 hp ->D2_2 hp | 50.0 | 37.5 | 25.3 | 25.4 | 94.9 | **99.9** |
| Task 1_4 | D2_0_1_2 hp ->D1_0 hp | 25.0 | 60.3 | 25.0 | 32.5 | 100 | **100** |
| Task 1_5 | D2_0_1_2 hp ->D1_1 hp | 25.0 | 60.3 | 25.0 | 32.5 | 100 | **100** |
| Task 1_6 | D2_0_1_2 hp ->D1_2 hp | 25.0 | 60.3 | 25.0 | 32.5 | 100 | **100** |
| Task 2_1 | D1_0_1_2 hp ->D3_0 hp | 25.0 | 50.0 | 50.0 | 50.0 | 100 | **100** |
| Task 2_2 | D1_0_1_2 hp ->D3_1 hp | 25.0 | 50.0 | 50.0 | 50.0 | 99.6 | **100** |
| Task 2_3 | D1_0_1_2 hp ->D3_2 hp | 25.0 | 50.0 | 50.0 | 50.0 | 75.1 | **98.9** |
| Task 2_4 | D3_0_1_2 hp ->D1_0 hp | 25.0 | 25.0 | 50.0 | 58.5 | 100 | **100** |
| Task 2_5 | D3_0_1_2 hp ->D1_1 hp | 25.0 | 25.0 | 50.0 | 58.5 | 100 | **100** |
| Task 2_6 | D3_0_1_2 hp ->D1_2 hp | 25.0 | 25.0 | 50.0 | 58.5 | 100 | **100** |

TABLE III
ABLATION EXPERIMENTS
(A=VAE_FEATURE, B=LABEL, C=OT, D=SAE, SAMPLE SIZE = 100)

| | ACC. (%) | ANOVA. | Params(m) | Runtime (s) |
|---|---|---|---|---|
| A | 25.0 | —— | —— | —— |
| A+B | 25.0 | —— | —— | —— |
| A+C | 81.6 | 0.061 | —— | —— |
| A+B+C | 98.2 | 0.008 | 0.49 | 40 |
| D+B+C | 99.9 | 0.019 | 0.48 | 32 |

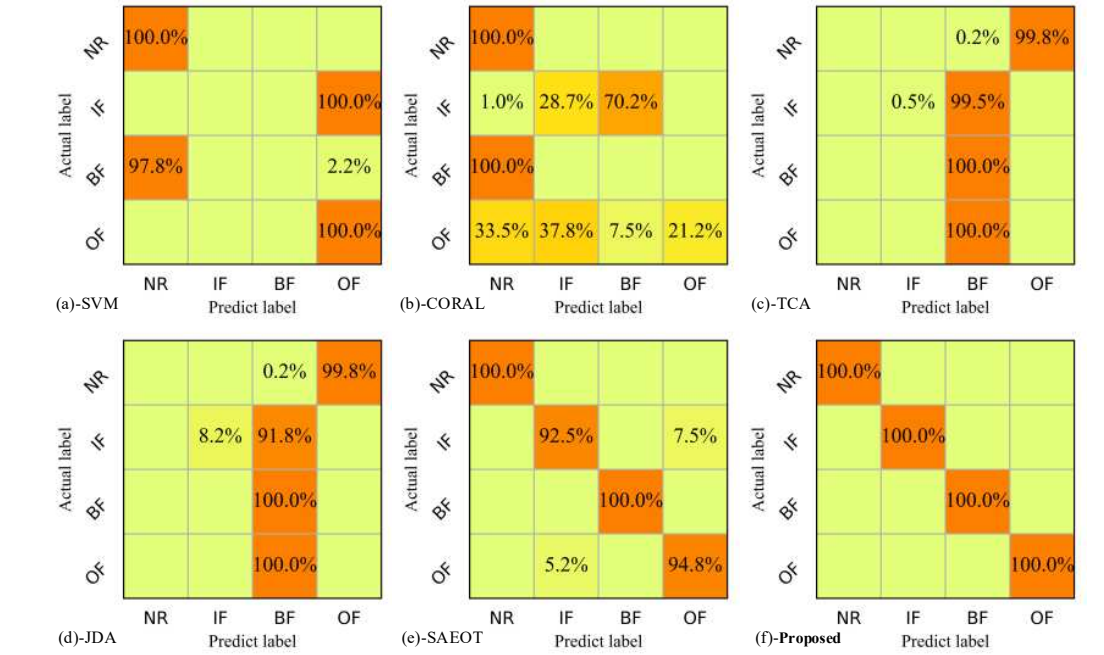



Fig. 9. Confusion matrix for each method of Task 1_1.

## *B. Compared Approaches*

For comparison purposes, several machine learning techniques are deployed in the experiments of this paper, including support vector machine (SVM), correlation alignment (CORAL), transfer component analysis (TCA), and joint distribution adaptation (JDA). In addition, a stacked auto-encoder network is used instead of VAE network for comparison experiments in this paper, and this method is noted as SAEOT.

The laboratory settings and detailed descriptions of these alternatives are as follows: for all adjustable parameters of SVM, CORAL, we use the default values suggested by the Pycharm software. Among them, CORAL's classification algorithm uses the k-Nearest Neighbor (KNN). The Gaussian kernel functions of TCA and JDA are chosen to run in MATLAB environment. Additionally, the average maximum accuracy is taken under different kernel function widths (1, 2, 4, 8) and regularization parameters (0.1, 0.01, 0.001). SAEOT and MDVAEOT choose their respective optimal parameters (e.g., learning rate, number of iterations, activation function) except for their own encoder structure.

We also performed ablation experiments shown in Table III. “A” indicates that VAE is used as the feature extractor. “A+B”

denotes consideration of joint distribution of features and labels with VAE as the feature extractor. “C” represents the inclusion of the optimal transport distance. “D” indicates that SAE is used as the feature extractor. Since the last two methods in Table III contain the same section that references the OT library, we only count the number of parameters for these two methods in addition to the OT section. We use Task 2 (sample size = 100) as an example of an ablation experiment to verify the validity of each part of the proposed method

### C. Experimental Results and Analysis

#### 1) Cross-machine fault diagnosis results

The accuracies of 6 methods in 12 cross-machine situations are shown in Tables II. It can be seen that the average accuracy of the proposed method is much higher than that of the other five compared methods. It is clear that the traditional intelligent algorithms and domain adaptation methods are not effective for cross-machine diagnosis. Even in Task 1, where the effectiveness of TCA with JDA is lower than that of the traditional method without domain adaptation. This is caused by the excessive differences in data distribution across machines and the absence of a better common subspace to project features.

The confusion matrix for each method in the Pycharm environment for Task 1_1 (i.e., D1_0_1_2 hp -> D2_0 hp) is shown in Fig. 9, panels (a) to (f) represent the confusion matrix plots for SVM, CORAL, TCA, JDA, SAEOT and the suggested solution, respectively.From the classification details of the confusion matrix diagrams, it can be seen that the first four methods are less effective. Additionally, CORAL, TCA, and JDA can basically classify only one state accurately. Combined with Table III, although SAEOT can perform better than the four methods mentioned above and basically achieves the accuracy of the proposed method, its stability is inferior to that of the suggested approach.

#### 2) Visualization of learned features

To visualize the classification performances of the compared methods, we take Task1_1 task (i.e., D1_0_1_2 hp -> D2_0 hp) as an example and check the scatter plots of the first two features extracted. The visualization of feature differences using t-SNE [36] is shown in Figures 10 and 11. There are four categories in both domains, which are represented by four notations. All source samples are represented in green in Figure 10 and all target samples are represented in red. Figure 11 shows the features obtained using the last four transfer learning methods, with yellow and red representing samples in the

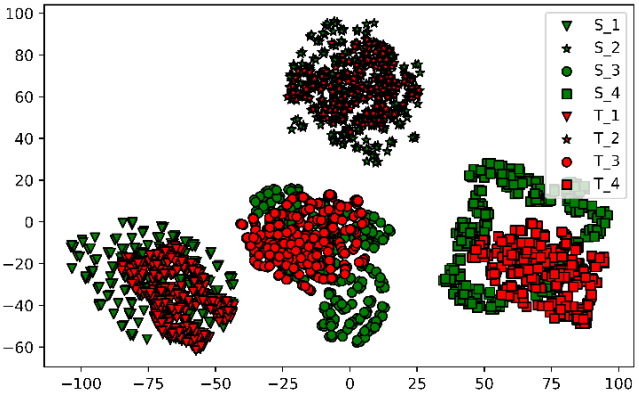



Fig. 10. Original feature distribution of Task1_1.

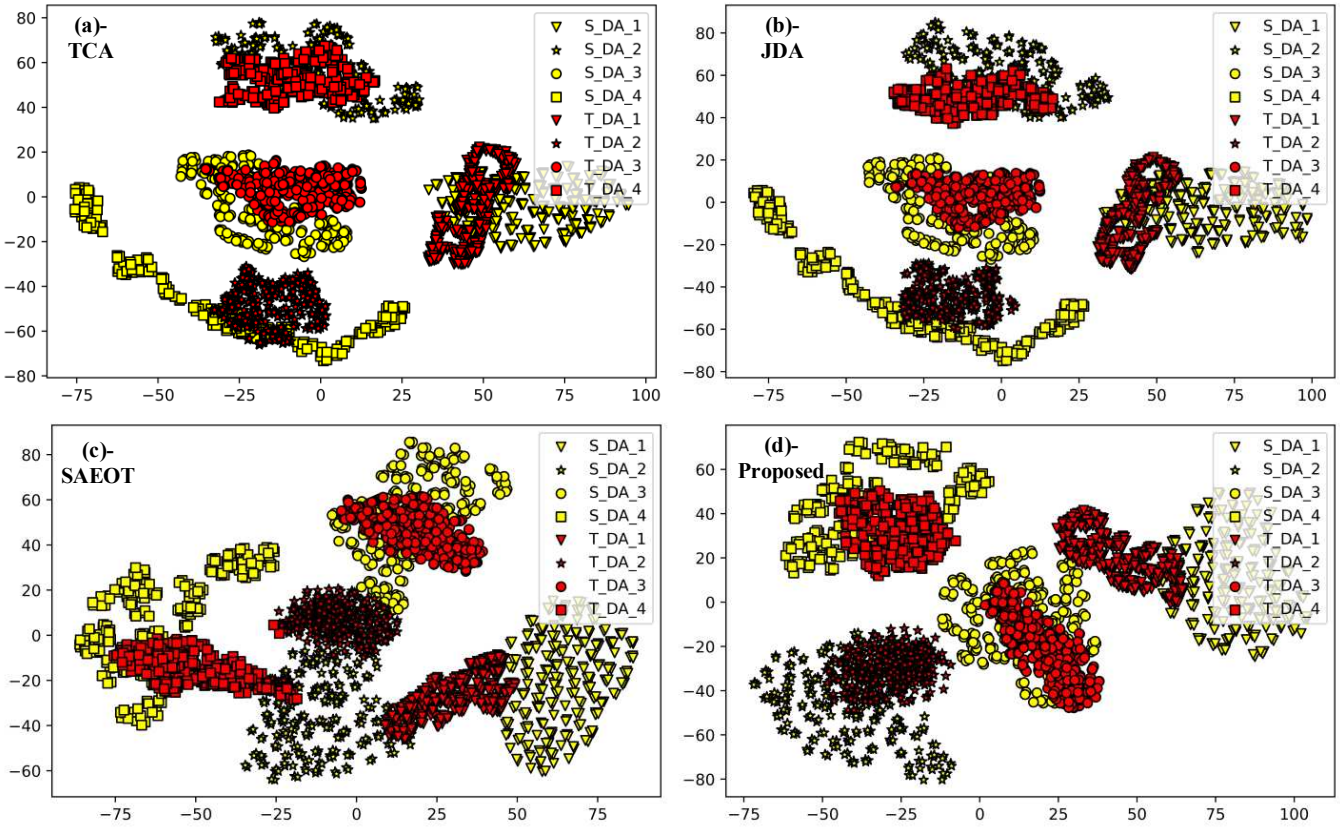



Fig. 11. Comparative analysis of the distributions of features of different methods for Task1_1.

source and target domains respectively. We expect samples of the same shape with different colours to be classified within a domain.

As shown in Figure 10, the gaps between the samples of the same category across machines are much smaller than that of the different categories of the same machine, which verifies fact 4. As shown in Figure 11, TCA and JDA cannot apply the discriminative knowledge of fault states from the source domain to the target domain effectively. SAEOT and the proposed method both achieve relatively good results. However, observing the accuracy of using SAEOT in Task 2, it is noticed that the accuracy sometimes appears to be around 75%. Such low accuracy has occurred in every subtask of the training task. This indicates that the diagnostic algorithm of SAEOT is not stable enough. Such a phenomenon can be explained by the distribution of features obtained by processing the data with SAE and VAE.

As can be seen from the feature distribution in Figure 11(c), the SAE-based training makes the features tend to be tightly clustered in the center and gradually become sparse and scattered. The ultimate goal in reducing the dimensionality of the features is to retain the main structural information of the features in a simplified representation while reducing the number of dimensions. Due to the lack of regularity in the hidden space of SAE, such an approach makes the feature distribution between classes prone to overlap and reduces the accuracy of the classification after training. For the VAE approach, the feature visualization results are revealed in Figure 11(d). In comparison with SAE, VAE has a more uniform distribution in the hidden variable space. This result can more clearly delineate feature distribution between categories, and helps to improve the accuracy and stability of the classification. Such an advantage is attributed to the presupposition of the probability distribution of the hidden variables in the VAE method.

*3) Hyperparametric analysis*

We analyzed the effect due to different numbers of source samples and different learning rates on the accuracy, which is shown in Figures 12 and 13. The quantity of source samples under each working condition is selected to 100, 200, 300, and 400. Additionally, various learning rates of 0.001, 0.0015, 0.002, 0.0025, 0.003, 0.004, and 0.0005 are set.

The outcomes reveal that the proposed method does not require a consistent sampling frequency (e.g., 12kHz to 48kHz) for cross-machine fault diagnosis. The accuracy analysis in Fig. 12 and the corresponding ANOVA in Fig. 13 show that the proposed method maintains good accuracy and stability at a learning rate of 0.002. And the source dataset of the method only needs to mix a sample set of about 6000 sample points (sampling frequency from 12kHz to 48kHz, sample size = 100) for each working condition to achieve good results. Therefore, it is not necessary for the method to collect numerous samples for a single working condition in the source domain. Combined with Table III, it only takes about 40s to complete such an end-to-end process from acquiring the sampled data to diagnosing the result. The advantage of small computational load of proposed method is greatly beneficial for industrial applications.

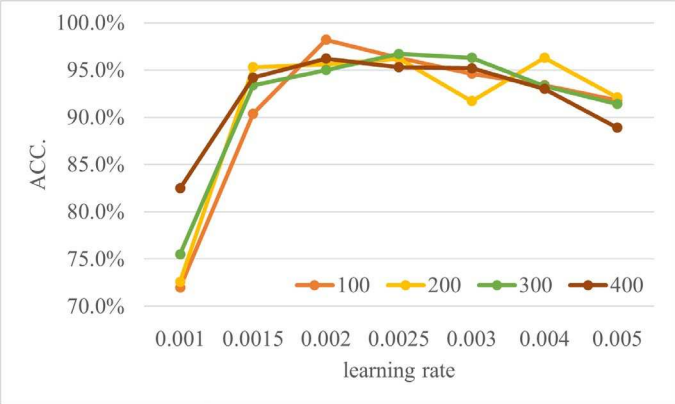



Fig. 12. Accuracies of various learning rates and sample sizes for Task2.

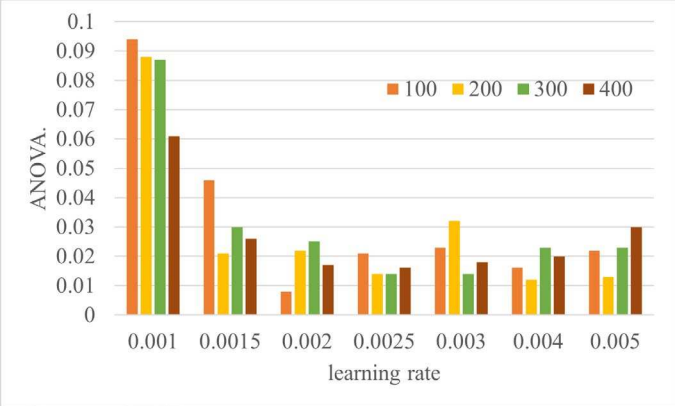



Fig. 13. ANOVA of accuracies for various learning rates with sample sizes for Task2.

All experiments in this paper were conducted on an Intel i5-10400 CPU, but more capable CPU or GPU could shrink the algorithm convergence time.

## V. Conclusions

In this article, we investigate the cross-machine fault diagnosis problem of rolling bearings, and propose a new variational auto-encoder based multi-source deep domain model using optimal transport for cross-machine diagnosis. It is radically dissimilar to most diagnostic models where both train and test data belong to the same machine.

The method enables end-to-end cross-machine fault class diagnosis using multi-source data from a labeled machine and unlabeled target data (from another machine to be diagnosed). Furthermore, the OT distance measures the difference of the both domain features, allowing source fault features to be shared to the target domain. The experimental outcomes confirm the superiority of VAE over SAE by t-SNE visualization. The impact of two hyperparameters (learning rate and sample size) on the accuracy of the algorithm is also examined. Experiments show that the suggested approach outperforms the main existing ones in terms of computational speed, accuracy and practicality, and is more suitable for industrial applications.

## REFERENCES


[1] W. Ma, Y. Zhang, L. Ma, R. Liu, and S. Yan, 'An unsupervised domain adaptation approach with enhanced transferability and discriminability for bearing fault diagnosis under few-shot samples', *Expert Systems with Applications*, vol. 225, p. 120084, Sep. 2023, doi: 10.1016/j.eswa.2023.120084.

[2] C. Ma, Y. Li, X. Wang, and Z. Cai, 'Early fault diagnosis of rotating machinery based on composite zoom permutation entropy', *Reliability Engineering & System Safety*, vol. 230, p. 108967, Feb. 2023, doi: 10.1016/j.ress.2022.108967.

[3] Z.-H. Liu, L.-B. Jiang, H.-L. Wei, L. Chen, and X.-H. Li, 'Optimal Transport Based Deep Domain Adaptation Approach for Fault Diagnosis of Rotating Machine', *IEEE Trans. Instrum. Meas.*, pp. 1–1, 2021, doi: 10.1109/TIM.2021.3050173.

[4] J. Harmouche, C. Delpha, and D. Diallo, 'Improved Fault Diagnosis of Ball Bearings Based on the Global Spectrum of Vibration Signals', *IEEE Trans. Energy Convers.*, vol. 30, no. 1, pp. 376–383, Mar. 2015, doi: 10.1109/TEC.2014.2341620.

[5] M. Sadoughi and C. Hu, 'Physics-Based Convolutional Neural Network for Fault Diagnosis of Rolling Element Bearings', *IEEE Sensors J.*, vol. 19, no. 11, pp. 4181–4192, Jun. 2019, doi: 10.1109/JSEN.2019.2898634.

[6] T. Berghout, L.-H. Mouss, T. Bentrcia, and M. Benbouzid, 'A Semi-Supervised Deep Transfer Learning Approach for Rolling-Element Bearing Remaining Useful Life Prediction', *IEEE Trans. Energy Convers.*, vol. 37, no. 2, pp. 1200–1210, Jun. 2022, doi: 10.1109/TEC.2021.3116423.

[7] K. N. Ravikumar, A. Yadav, H. Kumar, K. V. Gangadharan, and A. V. Narasimhadhan, 'Gearbox fault diagnosis based on Multi-Scale deep residual learning and stacked LSTM model', *Measurement*, vol. 186, p. 110099, Dec. 2021, doi: 10.1016/j.measurement.2021.110099.

[8] D. Peng, H. Wang, Z. Liu, W. Zhang, M. J. Zuo, and J. Chen, 'Multibranch and Multiscale CNN for Fault Diagnosis of Wheelset Bearings Under Strong Noise and Variable Load Condition', *IEEE Trans. Ind. Inf.*, vol. 16, no. 7, pp. 4949–4960, Jul. 2020, doi: 10.1109/TII.2020.2967557.

[9] Z. Tang, M. Wang, T. Ouyang, and F. Che, 'A wind turbine bearing fault diagnosis method based on fused depth features in time–frequency domain', *Energy Reports*, vol. 8, pp. 12727–12739, Nov. 2022, doi: 10.1016/j.egyr.2022.09.113.

[10] Y. Hou *et al.*, 'Acoustic feature enhancement in rolling bearing fault diagnosis using sparsity-oriented multipoint optimal minimum entropy deconvolution adjusted method', *Applied Acoustics*, vol. 201, p. 109105, Dec. 2022, doi: 10.1016/j.apacoust.2022.109105.

[11] J. Yang, J. Liu, J. Xie, C. Wang, and T. Ding, 'Conditional GAN and 2-D CNN for Bearing Fault Diagnosis With Small Samples', *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021, doi: 10.1109/TIM.2021.3119135.

[12] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang, 'Deep Model Based Domain Adaptation for Fault Diagnosis', *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, Mar. 2017, doi: 10.1109/TIE.2016.2627020.

[13] S. J. Pan and Q. Yang, 'A Survey on Transfer Learning', *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[14] V. Singh and N. K. Verma, 'Intelligent Condition-Based Monitoring Techniques for Bearing Fault Diagnosis', *IEEE Sensors J.*, vol. 21, no. 14, pp. 15448–15457, Jul. 2021, doi: 10.1109/JSEN.2020.3021918.

[15] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, 'Domain Adaptation via Transfer Component Analysis', *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011, doi: 10.1109/TNN.2010.2091281.

[16] Z.-H. Liu, B.-L. Lu, H.-L. Wei, X.-H. Li, and L. Chen, 'Fault Diagnosis for Electromechanical Drivetrains Using a Joint Distribution Optimal Deep Domain Adaptation Approach', *IEEE Sensors J.*, vol. 19, no. 24, pp. 12261–12270, Dec. 2019, doi: 10.1109/JSEN.2019.2939360.

[17] S. Yang, X. Kong, Q. Wang, Z. Li, H. Cheng, and K. Xu, 'Deep multiple auto-encoder with attention mechanism network: A dynamic domain adaptation method for rotary machine fault diagnosis under different working conditions', *Knowledge-Based Systems*, vol. 249, p. 108639, Aug. 2022, doi: 10.1016/j.knosys.2022.108639.

[18] J. Li, R. Huang, G. He, Y. Liao, Z. Wang, and W. Li, 'A Two-Stage Transfer Adversarial Network for Intelligent Fault Diagnosis of Rotating Machinery With Multiple New Faults', *IEEE/ASME Trans. Mechatron.*, vol. 26, no. 3, pp. 1591–1601, Jun. 2021, doi: 10.1109/TMECH.2020.3025615.

[19] Q. Hu, X. Si, A. Qin, Y. Lv, and M. Liu, 'Balanced Adaptation Regularization Based Transfer Learning for Unsupervised Cross-Domain Fault Diagnosis', *IEEE Sensors J.*, vol. 22, no. 12, pp. 12139–12151, Jun. 2022, doi: 10.1109/JSEN.2022.3174396.

[20] J. Chen, J. Wang, J. Zhu, T. H. Lee, and C. W. de Silva, 'Unsupervised Cross-Domain Fault Diagnosis Using Feature Representation Alignment Networks for Rotating Machinery', *IEEE/ASME Trans. Mechatron.*, vol. 26, no. 5, pp. 2770–2781, Oct. 2021, doi: 10.1109/TMECH.2020.3046277.

[21] Z. Chai, C. Zhao, and B. Huang, 'Multisource-Refined Transfer Network for Industrial Fault Diagnosis Under Domain and Category Inconsistencies', *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9784–9796, Sep. 2022, doi: 10.1109/TCYB.2021.3067786.

[22] B. Yang, Y. Lei, S. Xu, and C.-G. Lee, 'An Optimal Transport-Embedded Similarity Measure for Diagnostic Knowledge Transferability Analytics Across Machines', *IEEE Trans. Ind. Electron.*, vol. 69, no. 7, pp. 7372–7382, Jul. 2022, doi: 10.1109/TIE.2021.3095804.

[23] C. Shen, Y. Xia, X. Jiang, Z. Chen, L. Kong, and Z. Zhu, 'Optimal Transport-Based Multisource Student Teacher Learning Network for Bearing Fault Diagnosis Under Variable Working Conditions', *IEEE Sensors J.*, vol. 22, no. 16, pp. 16392–16401, Aug. 2022, doi: 10.1109/JSEN.2022.3190513.

[24] X. Li, X.-D. Jia, W. Zhang, H. Ma, Z. Luo, and X. Li, 'Intelligent cross-machine fault diagnosis approach with deep auto-encoder and domain adaptation', *Neurocomputing*, vol. 383, pp. 235–247, Mar. 2020, doi: 10.1016/j.neucom.2019.12.033.

[25] Y. Zhou, Y. Dong, H. Zhou, and G. Tang, 'Deep Dynamic Adaptive Transfer Network for Rolling Bearing Fault Diagnosis With Considering Cross-Machine Instance', *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021, doi: 10.1109/TIM.2021.3112800.

[26] S. Wan, J. Liu, X. Li, Y. Zhang, K. Yan, and J. Hong, 'Transfer-learning-based bearing fault diagnosis between different machines: A multi-level adaptation network based on layered decoding and attention mechanism', *Measurement*, vol. 203, p. 111996, Nov. 2022, doi: 10.1016/j.measurement.2022.111996.

[27] D. P. Kingma and M. Welling, 'Auto-Encoding Variational Bayes'. arXiv, Dec. 10, 2022. Accessed: Dec. 16, 2022. [Online]. Available: http://arxiv.org/abs/1312.6114

[28] B. Quanz, J. Huan, and M. Mishra, 'Knowledge transfer with low-quality data: A feature extraction issue', in *2011 IEEE 27th International Conference on Data Engineering*, Hannover, Germany: IEEE, Apr. 2011, pp. 769–779. doi: 10.1109/ICDE.2011.5767917.

[29] B. Yang, S. Xu, Y. Lei, C.-G. Lee, E. Stewart, and C. Roberts, 'Multi-source transfer learning network to complement knowledge for intelligent diagnosis of machines with unseen faults', *Mechanical Systems and Signal Processing*, vol. 162, p. 108095, Jan. 2022, doi: 10.1016/j.ymssp.2021.108095.

[30] A. K. Sharma and N. K. Verma, 'Quick Learning Mechanism With Cross-Domain Adaptation for Intelligent Fault Diagnosis', *IEEE Trans. Artif. Intell.*, vol. 3, no. 3, pp. 381–390, Jun. 2022, doi: 10.1109/TAI.2021.3123935.

[31] C. Shui, Q. Chen, J. Wen, F. Zhou, C. Gagné, and B. Wang, 'Beyond H-Divergence: Domain Adaptation Theory With Jensen-Shannon Divergence'.

[32] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, vol. 87. in Progress in Nonlinear Differential Equations and Their Applications, vol. 87. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-20828-2.

[33] Y. Xie, X. Wang, R. Wang, and H. Zha, 'A Fast Proximal Point Method for Computing Exact Wasserstein Distance', in *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, PMLR, Aug. 2020, pp. 433–453. Accessed: Dec. 16, 2022. [Online]. Available: https://proceedings.mlr.press/v115/xie20b.html

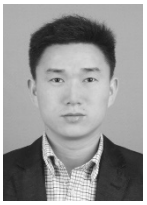
[34] N. Courty, R. Flamary, A. Habrard, and A. Rakotomamonjy, 'Joint distribution optimal transportation for domain adaptation', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Dec. 16, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/0070d23b06b1486a538c0eaa45dd167a-Abstract.html

[35] 'Bearing Data Center | Case School of Engineering | Case Western Reserve University'. https://engineering.case.edu/bearingdatacenter (accessed Dec. 16, 2022).

[36] P. Hajibabaee, F. Pourkamali-Anaraki, and M. A. Hariri-Ardebili, 'An Empirical Evaluation of the t-SNE Algorithm for Data Visualization in Structural Engineering', in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Pasadena, CA, USA: IEEE, Dec. 2021, pp. 1674–1680. doi: 10.1109/ICMLA52953.2021.00267.

**Shi-Zheng Yuan** received the B.Eng. degree in automation from the Hunan university of science and technology, Xiangtan, China, in 2020. He is currently pursuing the M.Sc. degree in automatic control and electrical engineering, at Hunan university of science and technology, Xiangtan, China.

His current research interests include deep learning algorithm design and fault diagnosis of wind turbine transmission chains.



**Zhao-Hua Liu** (M'16, SM'2022) received M.Sc. degree in computer science and engineering, and the PhD. degree in automatic control and electrical engineering from the Hunan University, China, in 2010 and 2012, respectively. He worked as a visiting researcher in the Department of Automatic Control and Systems Engineering at the University of Sheffield, United Kingdom, from 2015 to 2016.

He is currently a Professor of Automatic Control and Systems with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His current research interests include intelligent information processing and control of wind turbines, computational intelligence and learning algorithms design, machine learning aided fault diagnosis and prognosis, parameter estimation and intelligent control of permanent-magnet synchronous machine drives, and fault diagnosis based intelligent operations and maintenance for electric power equipment.

Dr. Liu has published more than 50 research papers in refereed journals and conferences, including IEEE TRANSACTIONS / JOURNAL / MAGAZINE. He is a regular reviewer for several international journals and conferences.

**Hua-Liang Wei** received the Ph.D. degree in automatic control from the University of Sheffield, Sheffield, U.K., in 2004.

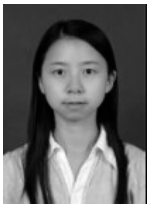
He is currently a senior lecturer with the Department of Automatic Control and Systems Engineering, the University of Sheffield, Sheffield, UK. His research focuses on evolutionary algorithms, identification and modelling for complex nonlinear systems, applications and developments of signal processing, system identification and data modelling to control engineering.

**Lei Chen** received the Ph.D. degree in automatic control and electrical engineering from the Hunan University, China, in 2017.

He is currently a Lecturer with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His current research interests include deep learning, network representation learning, information security of industrial control system and big data analysis.

**Ming-Yang Lv** received the Ph.D. degree in control theory and engineering with Hunan University, Changsha, China, in 2020.

He is currently a Lecturer with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His research interests include chaos theory and application, and modeling for complex process industries based on machine learning and deep learning.



**Xiao-Hua Li** received the B.Eng. degree in computer science and engineering from the Hunan University of Science and Engineering, Yongzhou, China, in 2007 and the M.Sc. degree in computer science from Hunan University, Changsha, China, in 2010.

She is currently a Lecturer of computer science with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. Her research interest includes evolutionary computation and machine learning.