

Research Article

A Novel Fault Diagnosis Method for Denoising Autoencoder Assisted by Digital Twin

Wenan Cai ¹, Qianqian Zhang ², and Jie Cui³

¹School of Mechanical Engineering, Jinzhong University, Jinzhong 030619, China

²School of Automation and Software Engineering, Shanxi University, Taiyuan 030006, China

³School of Mechanical Engineering, North University of China, Taiyuan, Shanxi 030051, China

Correspondence should be addressed to Wenan Cai; caiwenan65@163.com

Received 14 May 2022; Accepted 5 July 2022; Published 21 July 2022

Academic Editor: Jie Liu

Copyright © 2022 Wenan Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Digital twin (DT) is an important method to realize intelligent manufacturing. Traditional data-based fault diagnosis methods such as fractional-order fault feature extraction methods require sufficient data to train a diagnosis model, which is unrealistic in a dynamically changing production process. The ultrahigh-fidelity DT model can generate fault state data similar to the actual system, providing a new paradigm for fault diagnosis. This paper proposes a novel digital twin-assisted fault diagnosis method for denoising autoencoder. First, in order to solve the problem of limited or unavailable fault state data for machines in dynamically variable production scenarios, a DT model of the machine is established. The model can simulate a dynamically changing production process, thereby generating data for different failure states. Second, a novel denoising autoencoder (NDAE) with Mish as the activation function is proposed and trained using the source domain data generated by DT. Finally, in order to verify the effectiveness and feasibility of the proposed method, the method is applied to a fault diagnosis example of a triplex pump, and the results show that the method can realize intelligent fault diagnosis when the fault state data are limited or unavailable.

1. Introduction

As intelligent manufacturing becomes increasingly automated, digitized, and intelligent, more attention is paid to manufacturing process reliability and safety [1–3]. Minor failures in the production process can cause irreparable damage. Therefore, fault diagnosis is an important aspect of intelligent manufacturing [4–6]. Among the current machine learning methods, support vector machines, decision trees, and fractional order have been successfully applied in the field of fault diagnosis [7, 8]. In recent years, deep learning methods such as Bayesian networks, long short-term memory, and convolutional neural networks have been very popular in fault diagnosis due to their powerful modeling and representation capabilities [9–12]. The methods mentioned above can significantly improve the accuracy and efficiency of fault diagnosis under certain conditions. However, in order to obtain high fault diagnosis accuracy for deep learning methods, the primary condition

is that the source domain data should be sufficient and contain comprehensive fault diagnosis information. In practical industrial applications, machines are often in a dynamically changing production environment, and the health and fault information collected at this time are uncertain. Therefore, in the dynamically changing production process, it is difficult to collect a large amount of labeled fault data [13]. In addition, the machine is time-consuming and laborious to complete the degradation process, and the cost of marking a large amount of fault data is high. In order to prevent catastrophic accidents, many enterprises and factories do not allow machines to run to failure. Therefore, the above intelligent fault diagnosis methods are difficult to play a role in the dynamic changing production process [14].

In order to solve the above problems of insufficient training data and incomplete diagnostic information, some scholars have thought of transfer learning methods. This method can transfer a large amount of diagnostic information collected on a specific experimental platform to

dynamically changing production scenarios, solving the problem of insufficient training data [15]. For example, Zhang et al. proposed a first-layer wide convolutional deep neural network (WDCNN), the key of which is to use large convolution kernels in the first layer of convolution to extract short-term features. The convolution kernel parameters of the remaining convolutional layers except the first layer are reduced, which is conducive to deepening the network and suppressing overfitting [16]. Zhang et al. proposed a new CNN model, the advantage of which is that it does not require signal denoising preprocessing, which can realize fault diagnosis in noisy environments and variable working conditions [17]. Ren et al. proposed a new fault detection and classification method (DRCNN), which designed an important module “multiscale summation” for deep feature extraction. This method can combine features of multiple scales and different levels from unequal layers, which ensures the completeness of information [18]. However, the robustness of the diagnostic performance of transfer learning suffers in scenarios where working conditions and system characteristics are not fixed. Parameter transfer learning methods can assume that some parameters are shared between source tasks and target tasks, or the prior distribution of model hyperparameters is shared. Then, we use a small number of samples in the target domain to fine-tune the pretrained model, improve the overall performance of the model, and achieve a more robust fault diagnosis effect. However, the parameter transfer learning method also faces the problem of incomplete diagnostic information in the source domain [19].

With the rapid development and application of information technology, in recent years, digital twin (DT) technology has received more and more attention in various fields, such as product design and manufacturing, medical analysis, engineering construction, process optimization, and job shop scheduling. [20–22]. Also in the field of intelligent manufacturing, DT technology has also played a pivotal role and has become a powerful weapon to promote the development of intelligent manufacturing. It makes full use of physical model, sensor update, operation history, and other data, integrates multidisciplinary, multiphysical, multiscale, multiprobability simulation process, and completes the mapping in virtual space, thereby reflecting the full life cycle process of the corresponding physical equipment. DT technology can not only reduce design and maintenance costs but also improve manufacturing efficiency and quality. Ultrahigh-fidelity DT models can generate simulated data close to real systems, providing new opportunities for intelligent fault diagnosis. Wang et al. proposed a DT reference model for rotor system fault diagnosis. The requirements for building a digital twin model are discussed, and a model update scheme based on parameter sensitivity analysis is proposed to improve the adaptability of the model [23]. Jain et al. constructed a digital twin that can estimate the measurable characteristic output of a photovoltaic energy conversion unit (PVECU) in real time. A PVECU consists of a photovoltaic source and a source-level power converter [24]. Qin et al. proposed a full life cycle rolling bearing DT model driven by a combination of data and models. Through

an improved CycleGAN neural network, the simulated data in the virtual space are mapped to the physical space, and the results of the DT model are compared with the measured signals in the time and frequency domains to verify the effectiveness and feasibility of the proposed model [25]. Xu et al. proposed a two-stage DT fault diagnosis method (DFDD) based on deep transfer learning, which realizes fault diagnosis in the development and maintenance stages [26]. Qin et al. proposed a digital twin convolutional neural network model with multidomain input (DTCNNMI) in order to realize the misfire diagnosis of the diesel engine in a strong noise environment and different operating conditions [27]. However, most of the current research focuses on the conceptual model and key technologies of DT, and few people conduct more specific research on the fault diagnosis framework, mechanism, and algorithm to overcome the practical problem of limited diagnostic data.

In this paper, a novel digital twin-assisted fault diagnosis method for denoising autoencoder is proposed for the problem of machine intelligence fault diagnosis. The DT model can simulate a dynamically changing production process, thereby generating data of different fault states, and solving the problem of limited or unavailable fault state data for machines under dynamically variable production conditions. The main contributions of this paper are as follows:

- (1) A DT-assisted deep transfer learning fault diagnosis method is proposed, which is mainly used for fault diagnosis experiments of triplex pumps. A DT model of the machine is established to simulate the dynamically changing production process, thereby generating data for different failure states. The DT model is continuously updated during this process. The method solves the problem that the fault state data are limited or not used when the working state of the machine changes and the system characteristic changes.
- (2) A novel denoising autoencoder (NDAE) with Mish as the activation function is proposed, which has the properties of no upper bound, lower bound, smoothness, and nonmonotonicity compared with other activation functions.
- (3) A sparse penalty term is introduced to fully combine the advantages of sparse autoencoders and denoising autoencoders to effectively learn sparse feature representations from noisy samples.

2. Theoretical Background of Autoencoders

Convolutional neural network (CNN) is a commonly used network structure in deep learning methods and is currently widely used in the field of intelligent fault diagnosis of mechanical systems. However, the structure of CNN is relatively complex, and the amount of computation is relatively large compared with other deep learning methods. Compared with CNN, autoencoder has a simpler structure and stronger operability, which can train the model more easily and effectively. A type of neural network, after training, attempts to copy the input to the output. At

present, many improved forms have been derived from the autoencoder. On the basis of the autoencoder, the noise reduction autoencoder adds noise to the input data of the input layer in order to prevent the overfitting problem, so that the learned encoder is more robust. A sparse autoencoder is a special three-layer neural network with sparse constraints added to the general neural network. From the input layer to the hidden layer, the high-dimensional data are mapped to the low-dimensional data, and the projected low-dimensional data are restored to the original high-dimensional data from the hidden layer to the output layer [28]. Assuming that $x = [x_1, x_2, \dots, x_m]$ is a labeled m -dimensional real sample, the formula for noise sample $\tilde{x} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_m]$ is defined as follows:

$$\tilde{x} = x + N(0, \delta^2 \mathbf{I}), \quad (1)$$

where $N(0, \delta^2 \mathbf{I})$ represents Gaussian noise with noise level δ .

Then, the formulas of the feature vectors $\tilde{h} = [\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_n]$ and reconstruction vectors $\tilde{z} = [\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_m]$ of the noise samples are as follows:

$$\begin{aligned} \tilde{h} &= f_H(\mathbf{w}^{(1)}\tilde{x} + \mathbf{b}^{(1)}), \\ \tilde{z} &= f_O(\mathbf{w}^{(2)}\tilde{h} + \mathbf{b}^{(2)}), \end{aligned} \quad (2)$$

where f_H and f_O are the activation functions of the hidden layer and the output layer. $(\mathbf{w}^{(1)}, \mathbf{b}^{(1)})$ is the weights and biases of the input and hidden layers. Similarly, $(\mathbf{w}^{(2)}, \mathbf{b}^{(2)})$ is the weight and bias of the hidden layer and the output layer.

The formulas of the loss function l_1 , sparse penalty term l_2 , and weight decay term l_3 of MSE are expressed as follows:

$$\begin{aligned} l_1 &= \frac{1}{2} \sum_{i=1}^m (\tilde{z}_i - x_i)^2, \\ l_2 &= \beta \left(\sum_{j=1}^n r \log \frac{r}{\tilde{r}_j} + (1-r) \log \frac{1-r}{1-\tilde{r}_j} \right), \\ l_3 &= \frac{\lambda}{2} \left(\sum_{i=1}^m \sum_{j=1}^n \left((w_{ji}^{(1)})^2 + (w_{ji}^{(2)})^2 \right) \right), \end{aligned} \quad (3)$$

where β is the sparse penalty factor. r is the sparse constant. λ is the weight decay factor. $w_{ji}^{(1)}$ is the connection weight between the i th input unit and the j th hidden unit. Similarly, $w_{ji}^{(2)}$ is the connection between the j th hidden unit and the i th output unit connection weight.

Then, the overall loss function can be expressed as follows:

$$(L_S = l_1 + l_2 + l_3). \quad (4)$$

3. Proposed Method

3.1. NDAE Method. The NDAE method proposed in this paper is inspired by reference [28]. Based on the combination of sparse autoencoder and denoising autoencoder into sparse denoising autoencoder, a sparse penalty term is introduced, which can effectively learn the sparse features of

noise samples. In addition, inspired by reference [29], the Mish activation function with stronger learning ability is adopted. The ReLU activation function is the most widely used activation function in neural networks, and it mainly has the characteristics of having no upper bound and having a lower bound, which greatly limits its learning ability. Compared with ReLU, the Mish activation function has the characteristics of smoothness and nonmonotonicity. The smooth characteristics can make the network easier to optimize and improve the generalization performance, and the nonmonotonicity characteristics can improve the interpretability of the network. Comparison results on several datasets verify that Mish's metrics outperform ReLU and other activation functions for most tasks [29]. The waveform of the Mish function is shown in Figure 1. It can be seen from the figure that it allows a small negative gradient to flow in when it is negative, thereby ensuring information transfer and eliminating the dying ReLU phenomenon. The mathematical expression of the Mish activation function is as follows:

$$\begin{aligned} f_M(x) &= x \tanh(\text{softplus}(x)) \\ &= x \tanh(\ln(1 + e^x)), \\ \tanh(x) &= \frac{(e^x - e^{-x})}{(e^x + e^{-x})}, \end{aligned} \quad (5)$$

$$\text{softplus}(x) = \log(1 + e^x).$$

Mish is unbounded above and bounded below, and its first derivative can be defined as follows:

$$f'(x) = \frac{e^x \omega}{\delta^2}, \quad (6)$$

where $(\omega = 4(x+1) + 4e^{2x} + e^{3x} + e^x(4x+6))$, $(\delta = 2e^x + e^{2x} + 2)$.

The Mish activation function adds smoothness and nonmonotonicity to the ReLU activation function. These features can effectively retain the negative information of the data, make up for the deficiencies of ReLU, help information transfer, and have better expressiveness. It can be seen from formula (6) that the first derivative of the Mish activation function is differentiable; that is, the Mish activation function is continuously differentiable. This feature avoids singularities and thus avoids unwanted side effects when performing gradient-based optimization problems using the Mish activation function. In order to make the reconstructed output of the NDAE method infinitely close to the original input, a sigmoid activation function is selected at the output layer to normalize the input to the range of $[0, 1]$ into account. So the hidden and reconstructed outputs using the Mish activation function are

$$\begin{aligned} \tilde{h} &= f_M(\mathbf{w}^{(1)}\tilde{x} + \mathbf{b}^{(1)}), \\ \tilde{z} &= f_S(\mathbf{w}^{(2)}\tilde{h} + \mathbf{b}^{(2)}), \end{aligned} \quad (7)$$

where f_M and f_S are the Mish and sigmoid activation functions, respectively.

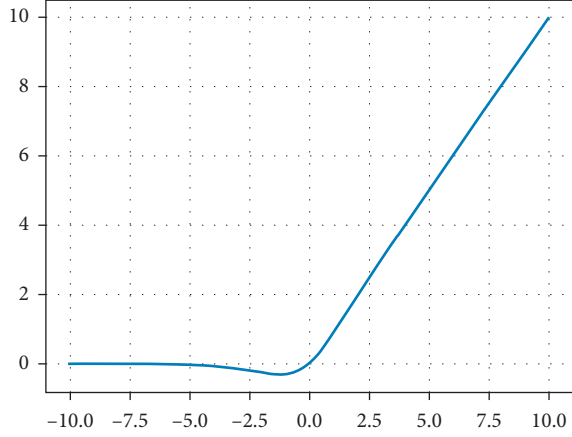


FIGURE 1: Mish activation function waveform.

To ensure that the real samples and the samples generated by the NDAE reconstruction are as similar as possible, the difference between the real samples and the reconstructed samples is reduced. To further measure the local similarity between the two, use the maximum correlation entropy instead of MSE, and use the gradient descent algorithm to adjust \mathbf{w} and \mathbf{b} , the correlation formula is as follows:

$$w_{q+1} = w_q - \xi_q \left(\frac{\partial L_N}{\partial w_q} \right) + \varepsilon (w_q - w_{q-1}),$$

$$b_{q+1} = b_q - \xi_q \left(\frac{\partial L_N}{\partial b_q} \right) + \varepsilon (b_q - b_{q-1}),$$
(8)

where q is the current number of iterations. L_N is the total loss function of the proposed method NDAE. ξ_q is the current learning rate. ε is the momentum factor. Then, the NDAE total loss function is as follows:

$$(L_N = -l_4 + l_2 + l_3),$$
(9)

where l_4 is the formula of maximum correlation entropy, which is more effective in local similarity measurement of complex signals than MSE. The formula for l_4 is as follows:

$$l_4 = \frac{1}{\sqrt{2\pi\tau}} \sum_{i=1}^m \exp\left(-\frac{(\tilde{z}_i - x_i)^2}{2\tau^2}\right),$$

$$\xi_{q+1} = \frac{\xi_q}{\rho},$$
(10)

where τ is the kernel width adjustment parameter. ρ is the decay factor. Using multiple NDAEs with softmax classifiers can be constructed to stack NDAEs to improve learning ability.

3.2. DT-Assisted NDAE Method. To solve the problem of limited or unavailable fault state data for machines under dynamically variable production conditions, a DT-assisted NDAE method is proposed in this paper. The overall framework of the method is shown in Figure 2. The green

part is the construction part of the DT model of the triplex pump. First, the simulation model of the real machine needs to be established, and then, the simulation model needs to be continuously updated to adapt to the dynamic and variable production environment. This paper updates the simulation model by minimizing the difference in system response between the simulation model and the measured data. The adaptively updated DT model is then used to simulate the fault state of the machine, generating comprehensive fault data required for fault diagnosis. The blue part in Figure 2 is the parameter transfer learning part of the new denoising autoencoder. First, the stacked NDAE model is constructed using the Mish activation function and maximum correlation entropy in 3.1, and then, a large amount of fault state data generated by the DT model are used as the training data in the source domain, which is input into the stacked NDAE model for pretraining. Finally, parameter transfer learning can greatly improve the training efficiency of stacked NDAE, so the parameter transfer learning method is used to realize machine fault diagnosis. It is worth noting that this paper selects a sample in the target domain to fine-tune the pre-trained stacked NDAE to further adjust the model parameters.

The shared parameters for parameter transfer learning in this paper are all hyperparameters, weights, and biases. It is worth noting that all weights and biases are pretrained before fine-tuning to ensure the effectiveness of parameter transfer learning.

4. Case Analysis

4.1. Experimental Description. In order to evaluate the effectiveness and feasibility of the proposed method, the method is applied to a fault diagnosis example of a triplex pump. The DT model of the triplex pump is shown in Figure 3. Inspired by reference [30], this paper imitates reference [30] and uses the Simscape module in Matlab to create a simulation model of a triplex pump. Triplex pumps have a crankshaft driving three plungers. Compared to single-piston pumps, one air chamber of the plunger is always vented, resulting in smoother flow and less pressure variation, thereby reducing material strain. The parameter values were then automatically tuned using Simulink design optimization so that the model produced results that matched the measured data to simulate the system behavior of a triplex pump in a dynamically variable production environment. Simulink design optimization selects parameter values for simulation, calculates the difference between the simulation curve and the measured curve to update the simulation model, and generates a simulation model with the system response function of modifying model parameters. Based on this difference, new parameter values are selected for a new simulation. The gradient of the parameter value is calculated to determine the direction in which the parameter should be adjusted. The DT model update of the triplex pump in this study is implemented through Simulink design optimization and automatically tunes the parameter values so that the model generates results that match the measured data.

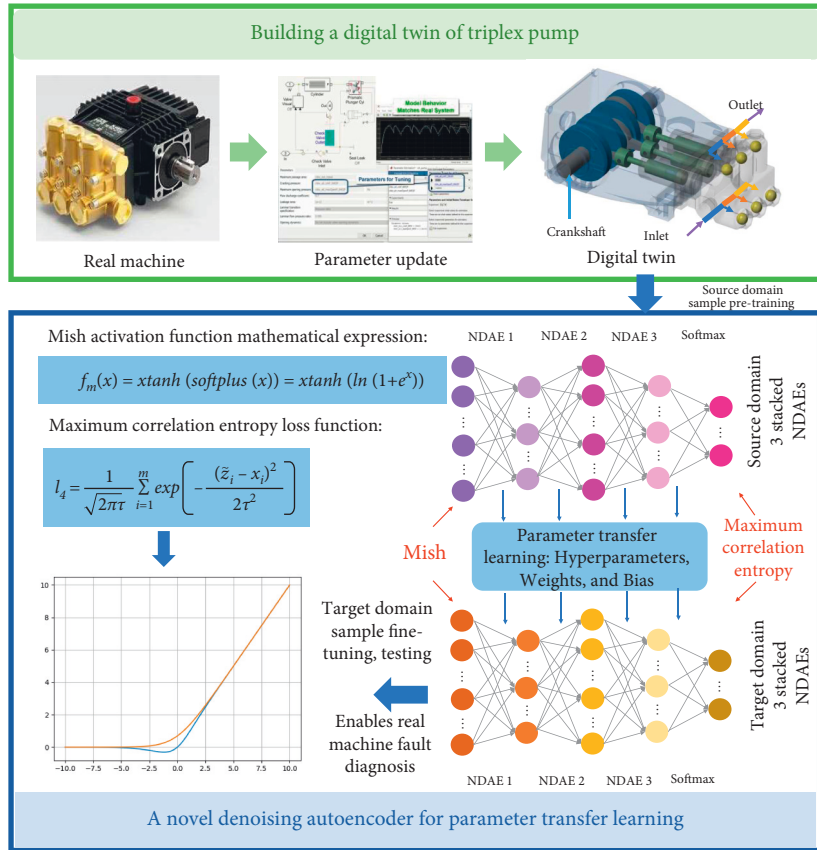


FIGURE 2: General framework of DT-assisted deep transfer learning method.

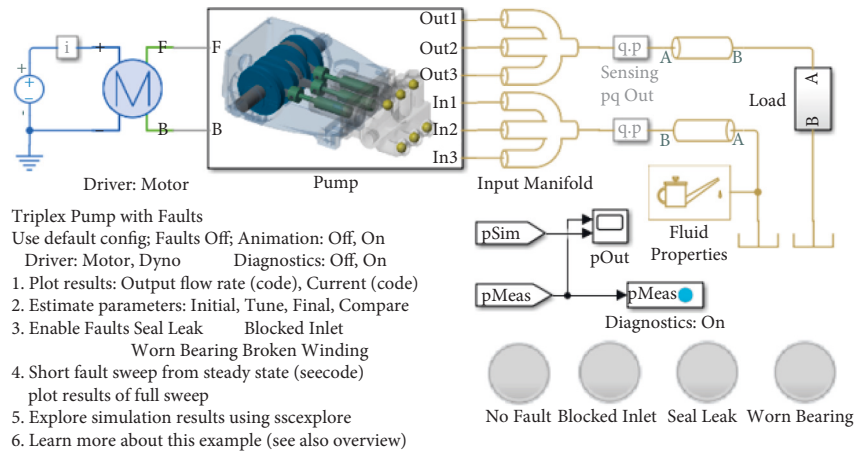


FIGURE 3: DT model of triplex pump.

The DT model of the triplex pump can simulate three typical pump failures, including seal leakage from the plunger, inlet blockage, and increased friction due to bearing wear. These fault conditions can be configured and toggled through the pump module dialog or commands. This paper collects data for seven fault states, including healthy state, three single faults, and three composite faults. Dataset A and dataset B are collected by simulating two scenarios. Dataset A is simulation data with original parameters. Dataset B is

simulation data collected when working conditions and system characteristics change, that is, actual situation data. Table 1 shows the detailed description of dataset A and dataset B, 125 samples are selected for each fault state, and each sample contains 1200 data points. Table 2 shows the detailed settings of the DT-assisted NDAE parameter transfer learning task. Table 3 shows the hyperparameter settings for stacked NDAE. The size of the first, second, and third hidden layers and other network structures are

TABLE 1: Seven working states of triplex pump.

Datasets	Working status	Number of samples	Labels
A/B	Healthy	125	1
	Seal leak	125	2
	Blocked inlet	125	3
	Bearing wear	125	4
	Seal leak and blocked inlet	125	5
	Seal leak and bearing wear	125	6
	Blocked inlet and bearing wear	125	7

TABLE 2: Detailed settings of parameter transfer tasks.

Method	Source domain training/test samples	Number of training/testing samples
Parameter transfer learning	Dataset A/B	75/50

TABLE 3: Hyperparameter settings for stacked NDAE methods.

Hyperparameters	Value	Hyperparameters	Value
The size of the first hidden layer	450	Kernel width	1.2
The size of the third hidden layer	100	Sparse penalty factor	5
Number of iterations	60	Initial learning rate	0.01
Weight decay coefficient	0.004	Decay factor	1.1
Noise level	0.08	Momentum	0.8

determined by experiments and experience. The number of iterations, initial learning rate, decay factor, and momentum are determined empirically, respectively. The selection of other hyperparameters is mainly based on reference [25].

4.2. Comparison Method. To verify the effectiveness and feasibility of the proposed method, it is compared with several state-of-the-art methods. Both the comparison method and the proposed method are tested using the fault data generated by the DT model.

- (1) SVM. The SVM algorithm is used to realize the fault classification of the triplex pump. SVM is a binary classification model that maps feature vectors to points in space, and its purpose is to find a line to better distinguish these points. Before the advent of deep learning, SVM was considered to be the better-performing algorithm in machine learning.
- (2) Stacked SDAE method. Stacked denoising autoencoders with ReLU as activation function.
- (3) LeNet-5 CNN. Fault classification of triplex pumps using a classic LeNet-5 convolutional neural network.

4.3. Experimental Results and Analysis. In this experiment, the effectiveness and feasibility of the proposed NDAE method are verified by the parameter transfer learning

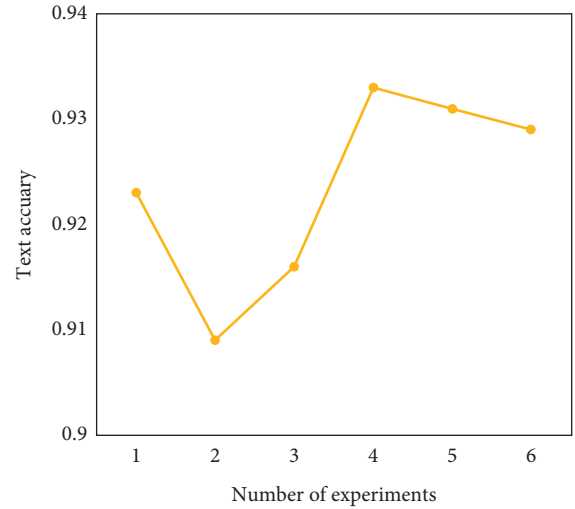


FIGURE 4: The fault diagnosis accuracy of the proposed method in ten experiments.

method. First, 75 samples are randomly selected from the 125 samples of dataset A as training samples in the source domain, and these 75 samples are input into the NDAE network for pretraining. Then, one sample is used from dataset B to fine-tune the pretrained network. This is because the DT model of the triplex pump can generate the data of the fault state, so just select a sample from the target domain to fine-tune the pretrained stacked NADE, and further construct the deep structure NADE to obtain better fault diagnosis results. Finally, 50 samples are used from dataset B for testing. In order to reduce the influence of random factors, the experiments were repeated six times; that is, six independent experiments were carried out using random samples for each method. The fault diagnosis accuracy of six experiments is shown in Figure 4. It can be seen from Figure 4 that the diagnostic accuracy of the six experiments exceeds 90%, and the average fault diagnosis accuracy is 92.4%.

To verify the effectiveness and feasibility of the proposed method, it is compared with SVM, stacked SDAE method, and LeNet-5 CNN method. To reduce the influence of random factors, the experiments were repeated six times; that is, six independent experiments were performed using random samples for each method. The experimental results are shown in Figure 5. The average accuracy of the SVM method, stacked SDAE method, and LeNet-5 CNN method is 76.5%, 87.8%, and 88.6%, respectively. It can be seen from the experimental results that the proposed method has higher fault diagnosis accuracy and is more conducive to the fault classification of the triplex pump.

5. Conclusion

In this paper, a DT-assisted NDAE parameter transfer learning fault diagnosis method is proposed, which is mainly used in the fault diagnosis experiment of the triplex pump.

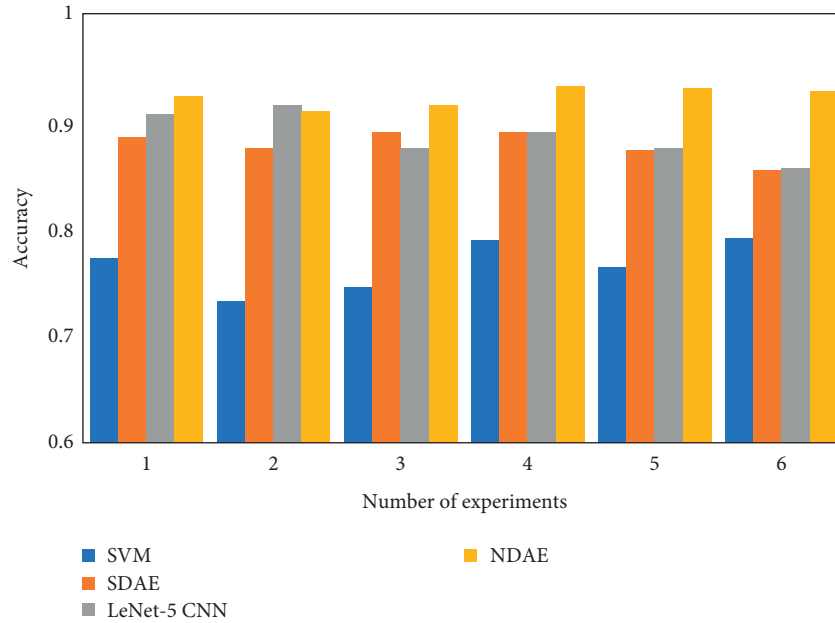


FIGURE 5: Fault diagnosis accuracy of four methods.

This method is designed to achieve high-accuracy fault classification when measurement data are insufficient or unavailable. We use the digital twin model of the machine to generate fault state data similar to the actual system to make up for the lack of data. In addition, the stacked autoencoder is improved, and the Mish activation function has no upper bound, lower bound, nonmonotonicity, and smoothness to increase the generalization performance of the network and the interpretability of the network. This ensures information transfer and eliminates the dying ReLU phenomenon. Finally, by generating simulation data of the triplex pump under various fault conditions, the effectiveness of fault diagnosis of the proposed NDAE method is verified. The results show that the ultrahigh-fidelity DT model can generate simulated data close to the real system, providing new opportunities for intelligent fault diagnosis. The DT-assisted NDAE parameter transfer learning fault diagnosis method can realize intelligent fault diagnosis of mechanical systems in dynamically changing production environments.

Although the DT-assisted NDAE parameter transfer learning fault diagnosis method can effectively improve the fault diagnosis accuracy of the model, the construction of the machine's DT model is a difficulty of this method. Therefore, the following research focuses on making full use of the main mechanism of DT to build DT models of other basic components such as bearings and combining deep transfer learning methods to improve fault diagnosis performance. How to further combine DT and deep transfer learning is the focus and difficulty of the next research.

Data Availability

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Wenan Cai wrote the paper; Qianqian Zhang revised the manuscript; Jie Cui revised the code of the manuscript. All authors have read and approved the final manuscript.

Acknowledgments

The authors would like to acknowledge Shanxi Province Scientific and technological innovation project of colleges and universities (2020L0606) and Shanxi Province Graduate Student Innovation Project (2021Y583).

References

- [1] A. Kusiak, "Smart manufacturing," *International Journal of Production Research*, vol. 56, no. 1-2, pp. 508–517, 2018.
- [2] S. Duan, W. Song, E. Zio, C. Cattani, and M. Li, "Product technical life prediction based on multi-modes and fractional Lévy stable motion," *Mechanical Systems and Signal Processing*, vol. 161, no. 5, Article ID 107974, 2021.
- [3] H. Liu, W. Song, Y. Zhang, and A. Kudreyko, "Generalized cauchy degradation model with long-range dependence and maximum lyapunov exponent for remaining useful life," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, Article ID 3512812, 2021.
- [4] 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," *Mechanical Systems and Signal Processing*, vol. 138, Article ID 106587, 2020.
- [5] Y. Li, X. Liang, Y. Yang, M. Xu, and W. Huang, "Early fault diagnosis of rotating machinery by combining differential

- rational spline-based LMD and K-L divergence,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 11, pp. 3077–3090, 2017.
- [6] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: a review,” *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, Aug. 2018.
- [7] Z. Yin and J. Hou, “Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes,” *Neurocomputing*, vol. 174, pp. 643–650, Jan. 2016.
- [8] H. Wang, J. Long, Z. Liu, F. You, and V. Ponomaryov, “Fault characteristic extraction by fractional lower-order bispectrum methods,” *Mathematical Problems in Engineering*, vol. 2020, Article ID 8823389, 24 pages, 2020.
- [9] L. Bennacer, Y. Amirat, A. Chibani, A. Mellouk, and L. Ciavaglia, “Self-diagnosis technique for virtual private networks combining bayesian networks and case-based reasoning,” *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 1, pp. 354–366, Jan. 2015.
- [10] X. He, Z. Wang, and Y. Li, “Joint decision-making of parallel machine scheduling restricted in job-machine release time and preventive maintenance with remaining useful life constraints,” *Reliability Engineering & System Safety*, vol. 222, 2022.
- [11] W. Zhao, Z. Wang, W. Cai et al., “Multiscale inverted residual convolutional neural network for intelligent diagnosis of bearings under variable load condition,” *Measurement*, vol. 188, Article ID 110511, 2022.
- [12] Z. Wang, J. Cui, W. Cai, and Y. Li, “Partial transfer learning of multidiscriminator deep weighted adversarial network in cross-machine fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.
- [13] J. Cui, Y. Li, Q. Zhang, Z. Wang, W. Du, and J. Wang, “Multi-layer adaptive convolutional neural network unsupervised domain adaptive bearing fault diagnosis method,” *Measurement Science and Technology*, vol. 33, no. 8, Article ID 085009, 2022.
- [14] Z. Wang, N. Yang, and N. Li, “A New Fault Diagnosis Method Based on Adaptive Spectrum Mode Extraction,” *Structural Health Monitoring*, vol. 20, 2021.
- [15] X. Li, Y. Hu, M. Li, and J. Zheng, “Fault diagnostics between different type of components: a transfer learning approach,” *Applied Soft Computing*, vol. 86, Article ID 105950, 2020.
- [16] W. Zhang, W. Ouyang, and L. Wen, “Collaborative and adversarial network for unsupervised domain adaptation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [17] Z. Pei, Z. Cao, M. Long, and J. Wang, “Multi-adversarial domain adaptation,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, February 2018.
- [18] Z. Wang, J. Zhou, Y. DuLei, W. Lei, and J. Wang, “Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution,” *Mechanical Systems and Signal Processing*, vol. 162, Article ID 108018, 2022.
- [19] R. Yan, F. Shen, C. Sun, and X. Chen, “Knowledge transfer for rotary machine fault diagnosis,” *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8374–8393, 2020.
- [20] F. Tao, F. Sui, A. Liu et al., “Digital twin-driven product design framework,” *International Journal of Production Research*, vol. 57, no. 12, pp. 3935–3953, 2019.
- [21] Y. Fang, C. Peng, P. Lou, Z. Zhou, J. Hu, and J. Yan, “Digital-twin-based job shop scheduling toward smart manufacturing,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 12, pp. 6425–6435, 2019.
- [22] Y. Yi, Y. Yan, X. Liu, Z. Ni, J. Feng, and J. Liu, “Digital twin-based smart assembly process design and application framework for complex products and its case study,” *Journal of Manufacturing Systems*, vol. 58, 2020.
- [23] J. Wang, L. Ye, R. X. Gao, C. Li, and L. Zhang, “Digital Twin for rotating machinery fault diagnosis in smart manufacturing,” *International Journal of Production Research*, vol. 57, no. 12, pp. 3920–3934, 2019.
- [24] P. Jain, J. Poon, J. P. Singh, C. Spanos, S. R. Sanders, and S. K. Panda, “A digital twin approach for fault diagnosis in distributed photovoltaic systems,” *IEEE Transactions on Power Electronics*, vol. 35, no. 1, pp. 940–956, 2020.
- [25] Y. Qin, X. Wu, and J. Luo, “Data-model combined driven digital twin of life-cycle rolling bearing,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1530–1540, 2022.
- [26] Y. Xu, Y. Sun, X. Liu, and Y. Zheng, “A digital-twin-assisted fault diagnosis using deep transfer learning,” *IEEE Access*, vol. 7, pp. 19990–19999, 2019.
- [27] C. Qin, Y. Jin, J. Tao, D. Xiao, H. Yu, and C. Liu, “DTCNNMI: A deep twin convolutional neural networks with multi-domain inputs for strongly noisy diesel engine misfire detection,” *Measurement*, vol. 180, 2021.
- [28] M. Xia, H. Shao, D. Williams, and L. Shu, “Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning,” *Reliability Engineering & System Safety*, vol. 215, 2021.
- [29] M. Diganta, “Mish: a self regularized non-monotonic activation function,” *Machine learning*, vol. 1, 2019.
- [30] S. Miller, “Predictive maintenance using a digital twin,” 2020, <https://uk.mathworks.com/company/newsletters/articles/predictive-maintenanceusing-a-digital-twin.html>.