



Mel Spectrogram-based advanced deep temporal clustering model with unsupervised data for fault diagnosis

Geonkyo Hong^{*}, Dongjun Suh^{*}

Department of Convergence & Fusion System Engineering, Kyungpook National University, Sangju 37224, South Korea



ARTICLE INFO

Keywords:

Anomaly detection
Data augmentation
Fault diagnosis
Mel spectrogram
Time series
Unsupervised learning

ABSTRACT

Fault diagnosis of mechanical equipment using data-driven machine learning methods has been developed recently as a promising technique for improving the reliability of industrial systems. However, these methods suffer from data sparsity due to the difficulty in data collection, which limits the feature extraction of anomalies. To solve this problem, we propose the mel spectrogram-based advanced deep temporal clustering (ADTC) model, which can extract and verify the features of unlabeled data through an unsupervised learning based autoencoder and the K-means. In addition, the ADTC model uses the proposed centroid based learning to obtain calibrated unsupervised learning data by minimizing the data point and target centroid distances for misclustered encoder output features in ensemble-based unsupervised learning. The classifier of the ADTC model uses a supervised learning based deep support vector machine network model, which is robust to nonlinear data, to diagnose the faults of the mechanical equipment. The proposed ADTC model was validated using mechanical equipment dataset with data augmentation to address the imbalanced dataset problem. During experiments, the mel spectrogram-based ADTC model exhibited the best performance in the various industrial environment with a prediction accuracy as high as 98.06%, outperforming other compared algorithms.

1. Introduction

The use of industrial systems to control mechanical equipment has increased manifold and has resulted in fresh complexities and uncertainties in the industrial environment. Consequently, the accurate diagnosis and prediction of faults in mechanical equipment is essential for performing effective maintenance and repairs to improve their reliability and productivity as well as reduce overall maintenance costs (Liu & Liu, 2003; Tung & Yang, 2009).

With recent advancements in artificial intelligence and big data in the field of fault diagnosis for efficient system utilization, interest in intelligent data-driven machine learning technology with high accuracy and easy access is increasing (Chalapathy & Chawla, 2019; Pang, Shen, Cao, Hengel, & Den., 2021). However, despite the promising performance reported in previous studies, the major shortcoming of data-driven fault diagnoses lies in the limitation of learning using a large amount of data (Shyu, Chen, & Iyengar, 2020). In industrial sites, only a limited amount of machine data can be collected, and abnormal datasets

with defects are more challenging to collect than normal datasets (Fahim & Sillitti, 2019). Because of imbalanced data, the feature extraction of anomalies is also limited.

Therefore, several studies have considered the sparseness of data in real environments by performing unsupervised learning-based failure diagnosis classification. (Wu, Zhang, Cheng, & Peng, 2021) performed machine turbine diagnostics with autoencoder (AE) models using a softmax classifier that reduces the reliance on unlabeled data. (Yang, Karimi, & Sun, 2021) implemented a wide kernel-based convolutional AE to learn features from raw signals important for diagnosing mechanical faults. (Liu et al., 2018) diagnosed faults by training the latent coding space on bearing training data using a categorical adversarial AE based on unsupervised learning. (Tao, Wang, Chen, Stojanovic, & Yang, 2020) generated spurious short time Fourier transform data from a CatGAN model and clustered the machine dataset for fault diagnosis. Although this study applied unsupervised learning to overcome the limitations of supervised learning when data are sparse, the accuracy was reduced when the model was applied to other domains. To date,

Peer review under responsibility of Submissions with the production note 'Please add the Reproducibility Badge for this item' the Badge and the following footnote to be added: The code (and data) in this article has been certified as Reproducible by the CodeOcean: <https://codeocean.com>. More information on the Reproducibility Badge Initiative is available at <https://www.elsevier.com/physicalsciencesandengineering/computerscience/journals>.

* Corresponding author.

E-mail addresses: gun2399@knu.ac.kr (G. Hong), dongjunsuh@knu.ac.kr (D. Suh).

existing studies have not considered multiple environmental conditions for mechanical equipment fault diagnosis and have focused on only one or two types of datasets corresponding to the mechanical equipment used for verification, limiting the adaptation of the results to other domains. When a data-driven model performs fault diagnosis, imbalance in the data can have a high impact on the performance of the model. Various studies have been conducted to analyze and solve the problems associated with imbalanced data (Johnson & Khoshgoftaar, 2019). (Shorten & Khoshgoftaar, 2019) reported that overfitting occurs during network training when the model learns a function with a very large variance because deep neural networks (DNNs) are strongly dependent on the data. Hence, deep learning frameworks that use data augmentation are required to improve the size and quality of the datasets. Chen and Jin (Chen & Jin, 2019) increased the performance accuracy of a model via data augmentation for highly imbalanced data. (Mikolajczyk & Grochowski, 2018) maximized model efficiency by applying various data augmentation methods to solve the imbalance issue. (Li, Li, Qu, & He, 2020) diagnosed gear defects using a deep sparse AE model with data augmentation. (Li, Li, & Ma, 2020) conducted a diagnostic study using data augmentation and unsupervised data exploration to reduce the effects of imbalanced data and sparsity on the generalization performance of an intelligent diagnostic model.

To overcome the data sparsity problem in this study, data augmentation is performed on a time-series-based vibration dataset, and unlabeled datasets are explored by combining the unsupervised learning-based autoencoder (AE) model and the K-means algorithm. In the unsupervised learning process, K-means is used to evaluate data features extracted from the encoder of the AE model. The K-means clustering makes it possible to know in advance how the data points extracted from the encoder of AE model affect the performance. By extracting data features through unsupervised learning, experiments can be performed that consider the real industrial environment, which has sparse data. This approach can use mel spectrogram image based unsupervised data to identify the characteristics of the fundamental structure of insufficient data, contributing to machine learning-based fault diagnosis using a mechanical equipment dataset. It also solves the domain adaptability problem by extracting unsupervised learning-based data that do not depend on a single target value from the sparse mechanical equipment dataset. Data points that are not included in the target centroid derived through K-means are weighted by training the proposed deep neural network (DNN) model using centroid-based learning so that they are moved closer to the target centroid. This method contributes to the performance of the final classifier model by obtaining high-quality features of the dataset. In the deep support vector machine network (DSVMN) classifier model, the output data of the encoder with added weights, which are not included in the target cluster, and the encoder output data, which are included in the target centroid obtained by K-means, are combined and used as the input value. On the basis of this algorithm, we propose the mel spectrogram image-based advanced deep temporal clustering (ADTC) model that can perform fault diagnosis using a class for each type of nonlinear-based mechanical equipment. The mel spectrogram-based image is used as the input value of the model to overcome the problem of combining a long length of data and complex frequency according to the sampling rate of the existing signal-based data in machine learning-based prediction. The image visualization method using the mel spectrogram, obtained by pre-processing the raw signals, performs signal processing on high-dimensional and complex raw audio signals, thereby scaling high and low frequencies to enable a representation of the datasets in the frequency domain. In addition, this method is widely used to improve performance in machine learning-based models for sound recognition through sound feature extraction (Tran & Lundgren, 2020; Arias-Vergara et al., 2021; Wang, Xue, Culhane, Diao, Ding, & Tarokh, 2020). Therefore, in this study, normal and abnormal data were identified using a representation in the frequency domain with the mel spectrogram-based mechanical equipment dataset. In validation experiments, the mel spectrogram image-

based model reflected various conditions of the industrial environment and diagnosed defects better than the raw signal-based model. The experimental results indicate that the proposed method provides superior performance in environments with various loads, noise signals, and lengths, thus indicating its application potential for industrial purposes. The contributions of this study are as follows.

- By verifying the unsupervised data extracted using the autoencoder through clustering, a high performance accuracy can be achieved for sparse mechanical equipment data.
- The proposed method can solve data sparsity and data imbalance problems using the DNN model to calibrate feature values for mechanical equipment that are not extracted well in unsupervised learning.
- The proposed method exhibits superior performance and shows a generalization of fault diagnosis for various mechanical equipment by overcoming data sparsity.
- To overcome the problems of data sparsity and imbalance in the limited datasets obtained in actual industrial sites, it is possible to increase the performance of the proposed ADTC model by augmenting the mechanical equipment dataset.
- The proposed approach solves the domain adaptability problem by extracting the features of the fundamental structure of the sparse mechanical equipment dataset using time-frequency based mel spectrogram image through unsupervised learning in advance and uses supervised learning to improve the performance of fault diagnosis.
- The robustness of the model was verified by demonstrating the superior performance of the proposed ADTC model through experiments varying the length, load, and levels of noise of various mechanical equipment datasets obtained in an industrial environment.

The remainder of this paper is organized as follows: Section 2 provides a detailed review of existing work, Section 3 presents and describes the proposed ADTC model. Section 4 outlines the experiments conducted and analyzes the results obtained. Finally, Section 5 summarizes our conclusions and suggests directions for future studies in this field.

2. Related work

Numerous studies have been conducted to assess and ensure the reliability and safety of data-driven fault diagnosis of mechanical equipment. (Rauber, da Silva Loca, & de Boldt, 2021) compared and analyzed the diagnostic performance of K-nearest neighbor, one-dimensional (1D) convolutional neural network (CNN), support vector machine (SVM), and random forest models using bearing data. (Brito, Susto, Brito, & Duarte, 2022) performed feature importance analysis through unsupervised learning-based machine fault diagnosis and Shapley additive explanations (SHAP) using frequency-based datasets to provide insight for each module. (Yang et al., 2021) performed feature extraction using an unsupervised learning-based deep belief network to diagnose faults in gearboxes and bearings, and partial least squares was used to optimize the supervised learning process to improve classification. (Saufi et al., 2020) diagnosed faults using a time-frequency based SSAE model with a limited number of samples to detect gearbox faults. (Varga, 2017) detected defects using a Large Memory Storage Retrieval model (LAMSTAR) based on signal processing-based short time fourier transform (STFT) data to detect the health status of bearings. (Sun, Yan, & Wen, 2018) applied nonlinear projection by compressing the data to identify defects in rotating machines. These data were used to perform fault diagnosis through the autoencoder. (Li, Zhang, & Ding, 2019) employed an attention mechanism to find information data segments and extract the identification function of the input to diagnose the defect of a rolling bearing. (Sohaib, Kim, & Kim, 2017) used a sparse stacked autoencoder (SAE)-based hybrid method to effectively diagnose bearing

faults of different severities. Although the above study performed data-based mechanical fault diagnosis, it did not consider various environments, and it has limitations because it did not test the adaptability of the method to other conditions such as noise and various loads in the actual industrial environment.

Many other studies have considered domain adaptability. (Xiao et al., 2021) diagnosed mechanical faults using the proposed the noisy domain adaptive marginal stacking denoising AE(NDAMsDA) model for noise domain adaptation. (Qian, Qin, Wang, & Liu, 2021) overcome a noisy environment with a convolutional AE, called CAE-DLTN, which uses Correlation Alignment (CORRE) to integrate domains. These studies addressed adaptability by performing fault diagnosis on noise, but there is a possibility of overfitting in a specific domain. (Li, Zhang, Ding, & Sun, 2019) minimized the multi-kernel maximum mean (MMD) of multiple layers for the learned representation in supervised learning to perform domain adaptability for bearing data. (Li, Hu, Zheng, Li, & Ma, 2021) evaluated the performance of domain adaptability for bearing data through central moment mismatching using a convolutional neural network-based model. (Chen, Zhao, He, Wei, & Yang, 2022) proposed a method for unsupervised learning-based bearing domain adaptability using the join sliced Wasserstein distance approach, which considers conditional probabilities. To overcome the limitations of previous studies, we consider the domain adaptation of industrial systems through fault diagnosis using mel spectrogram images reflecting various lengths, loads, and levels of noise of bearing and industrial machines.

3. Proposed method

Fig. 1 shows the proposed method for fault diagnosis of mechanical equipment. In this study, augmented mel spectrogram image-based learning data were used to overcome the data sparsity and data imbalance problems found in industry and machine learning-based methods. In addition, data sparseness can be overcome by extracting the features of the fundamental structure of the data in advance through unsupervised learning using these data. Diagnostic performance can further be improved by correcting the features of datasets that include inaccurate features using supervised learning.

3.1. Dataset formulation

Eq. (1) for the augmented mel spectrogram image-based input data $D_{us_{k=1}}$ of the feature extractor model based on the first unsupervised learning model ($k = 1$) is defined as n data samples as follows. Moreover, these input data $x_i^{us_{k=1}}$ in Eq. (2) can be expressed using $N_{us_{k=1}}$ dimensions as follows.

$$D_{us_{k=1}} = \{x_i^{us_{k=1}}\}_{i=1}^n \quad (1)$$

$$x_i^{us_{k=1}} \in R_{N_{us_{k=1}}} \quad (2)$$

The input data $D_{us_{k=2}}$ of the second unsupervised learning-based model ($k = 2$) for feature verification is defined as n data samples as follows in Eq. (3), where these data $x_i^{us_{k=2}}$ in Eq. (4) can be expressed using $N_{us_{k=2}}$ dimensions.

$$D_{us_{k=2}} = \{x_i^{us_{k=2}}\}_{i=1}^n \quad (3)$$

$$x_i^{us_{k=2}} \in R_{N_{us_{k=2}}} \quad (4)$$

The input data $D_{s_{k=1}}$ of the first supervised learning-based model ($k = 1$), which uses centroid learning in Eq. (5), is defined as n , where the input data in $x_i^{s_{k=1}}$ can be expressed as $N_{s_{k=1}}$ in Eq. (6).

$$D_{s_{k=1}} = \{x_i^{s_{k=1}}\}_{i=1}^n \quad (5)$$

$$x_i^{s_{k=1}} \in R_{N_{s_{k=1}}} \quad (6)$$

The second supervised learning model ($k = 2$) takes input data $D_{s_{k=2}}$ in Eq. (7), which consists of features extracted by the ensemble model, is represented by n data samples and $y_i^{s_{k=2}}$ represents the label corresponding to $x_i^{s_{k=2}}$. The input data $x_i^{s_{k=2}}$ in Eq. (8) of the supervised learning-based DSVMN classifier model can be represented using $N_{s_{k=2}}$ dimensions.

$$D_{s_{k=2}} = \{x_i^{s_{k=2}}, y_i^{s_{k=2}}\}_{i=1}^n \quad (7)$$

$$x_i^{s_{k=2}} \in R_{N_{s_{k=2}}} \quad (8)$$

In the experiment, a 4D image-based input is output in 2D through the encoder and used as input for the K-means based feature validator, DNN-based centroid based learning, and DSVMN classifier model. In addition, because the dimensions of the input of the unsupervised learning-based feature validator are the same as the dimensions of the output, the dimensions of the input of the supervised learning-based centroid learning model and classifier correspond. Therefore, the input dimensions of each model can be defined as Eq. (9). The same conditions for the training data and validation data of all mechanical equipment datasets were used for all data spaces $D_{us_{k=1}}$, $D_{us_{k=2}}$, $D_{s_{k=1}}$, and $D_{s_{k=2}}$.

$$N_{us_{k=1}} > N_{us_{k=2}} = N_{s_{k=1}} = N_{s_{k=2}} \quad (9)$$

3.2. Proposed method overview

In this section, we propose an ADTC model and mel spectrogram image-based algorithm that overcomes the problem of data sparsity and achieves the best performance on mechanical equipment datasets with various loads, noise signals, and time/length conditions. The detailed method of the proposed fault diagnosis is illustrated in Fig. 1 and Algorithm 1. The data were augmented to solve the imbalanced data problem in the time-series-based Case Western Reserve University (CWRU) bearing dataset and Malfunctioning Industrial Machine Investigation and Inspection (MIMII) dataset. The augmented datasets were converted into mel spectrogram images. Mel spectrogram image based unsupervised learning can solve the problem of data sparsity by extracting the features of the underlying structure, even with a small amount of data. It also helps address domain adaptability problems and improves performance by diagnosing faults using a data-driven machine-learning approach.

The mel spectrogram based proposed model can extract features from unsupervised data using the AE and K-means algorithm to overcome the data sparsity problem in real industry. In addition, we can generate high-quality data as feature vectors using several unsupervised learning models (Rajoub, 2020; Stevenson, Mues, & Bravo, 2021; Tavakoli & Heydarian, 2022; Zhang et al., 2022).

The encoder and decoder of the AE model are composed of 2D CNN and long short-term memory (LSTM) layers. In a 2D CNN, we extract visual patterns from mel spectrogram images and use LSTM to learn time series-based features. The AE is trained using normal data, and abnormal data values are predicted after learning the difference between the input value of the encoder and the reconstructed output value. Using this method, it is possible to solve the “curse of dimensionality” problem frequently encountered in deep learning by mapping input values into a low-dimensional space and learn the features of mechanical equipment data in an unsupervised manner. The low-dimensional points predicted by the learned encoder are used as input for the K-means algorithm, DNN, and DSVMN.

The K-means algorithm is used to validate unsupervised data using the output value of the encoder. Using the existing K-means algorithm to predict a mechanical equipment dataset with many classes and high levels of noise yields results with low accuracy. In this model, misclustered feature data reduce the performance. Therefore, the output value of the encoder corresponding to misclustered feature data is used as the input value of the DNN.

The DNN model with four fully connected layers based on the pro-

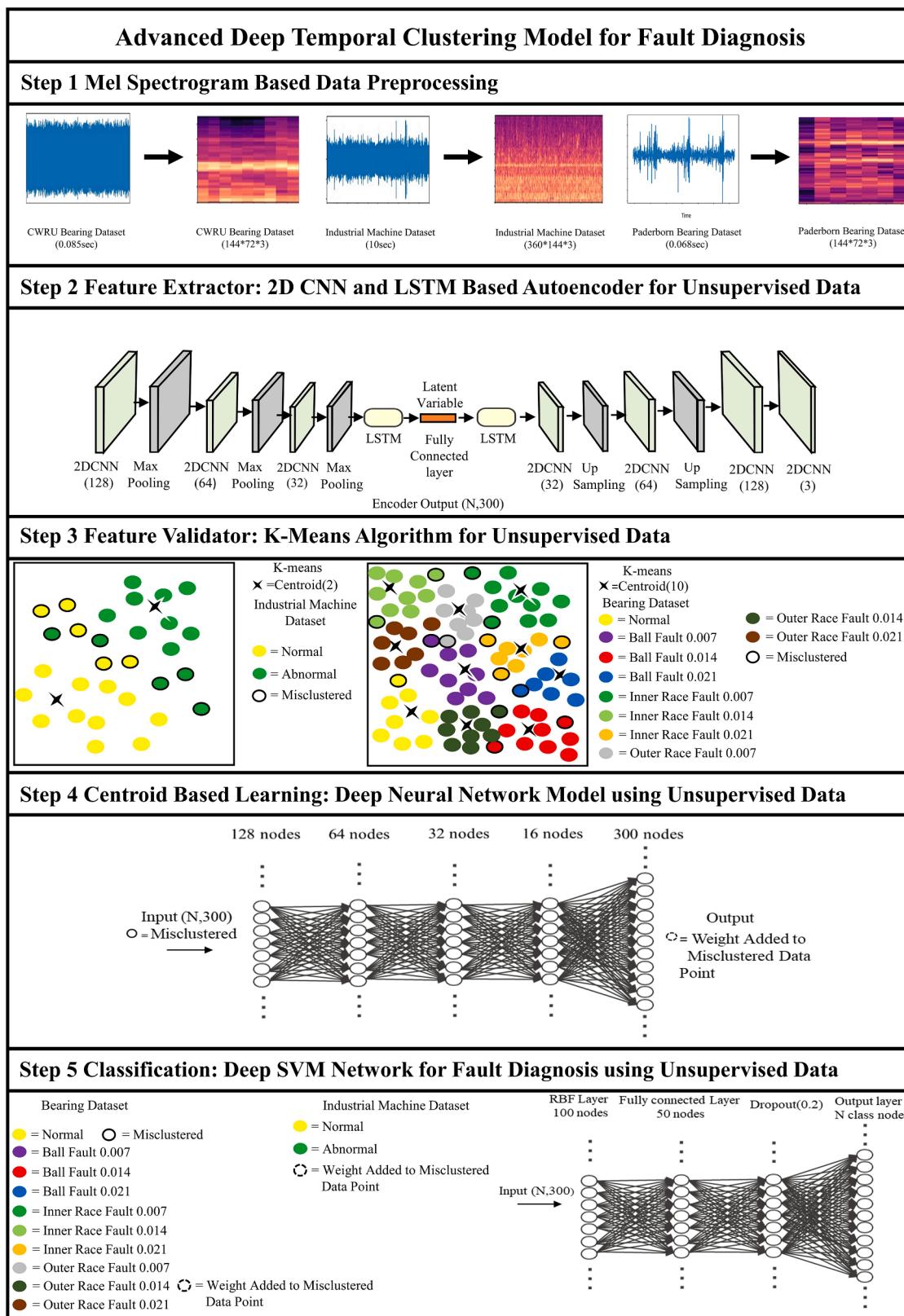


Fig. 1. Proposed method for mechanical equipment fault diagnosis.

posed Euclidean distance loss of the target centroid coordinate values is used to assign weights for each class to the machine equipment data points not included in the target cluster. By making the number of DNN models equal to the number of machine equipment classes, the target centroid distances are learned well.

Algorithm. 1. Proposed Method

Step 1: Input Value

Number of mel spectrogram image datasets with data augmentation

$$D_{us_{k-1}} = \{x_i^{us_{k-1}}\}_{i=1}^n, x_i^{us_{k-1}} \in R_{N_{us_{k-1}}}$$

Step 2: Unsupervised learning based on the feature extractor

Input image $\{x_i^{us_{k-1}}\}_{i=1}^n$ to the AE to extract the latent variable vector

$$D_{us_{k-2}} = \{x_i^{us_{k-2}}\}_{i=1}^n, x_i^{us_{k-2}} \in R_{N_{us_{k-2}}}$$

for epoch = 1,2,3,..., k do

Update θ_e, θ_{de} using Equations (12)–(17) and the data $\{x_i^{us_{k-1}}\}_{i=1}^n$
end for

Output: Predicted value $D_{us_{k-2}} = \{x_i^{us_{k-2}}\}_{i=1}^n, x_i^{us_{k-2}} \in R_{N_{us_{k-2}}}$ using the $\{x_i^{us_{k-1}}\}_{i=1}^n$ dataset

Step 3: Unsupervised learning based on the feature validator

Extract the feature value $ED_{(ij)}$ of the distance between the centroid μ_j and $\{x_i^{us_{k-2}}\}_{i=1}^n$ as the input for the K-means algorithm.

Initialize: centroids μ_j , number of data samples allocated to the center $\mu_j^{N_{assignment}}$, number of samples by class N_{class} ,

While True do:

for t = 1,2,3,..., n do

$$S_i^{(t)} = \{x_i^{us_{k-2}} : |x_i^{us_{k-2}} - \mu_j^{(t)}|^2 \leq |x_i^{us_{k-2}} - \mu_j^{(t)}|^2 \forall j, 1 \leq j \leq k\} \text{ using Equation (18)}$$

$$\mu_j^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_i^{us_{k-2}} \in S_i^{(t)}} x_i^{us_{k-2}} \text{ if } \mu_j^{(t)} = \mu_j^{(t+1)} \text{ do}$$

until Convergence

end for

if $\mu_j^{N_{assignment}} < N_{class} * 0.5$ then

Continue

else

Break

end for

Output: $D_{sk-1} = \{x_i^{sk-1}\}_{i=1}^n \leftarrow Misclustered(D_{us_{k-2}} = \{x_i^{us_{k-2}}\}_{i=1}^n)$,

Clustered($D_{us_{k-2}} = \{x_i^{us_{k-2}}\}_{i=1}^n$)

Step 4: Centroid based learning

In the DNN, the fault diagnosis of mechanical equipment is performed using the learned the mapping target centroid μ_j and the extracted feature $D_{sk-1} = \{x_i^{sk-1}\}_{i=1}^n \leftarrow Misclustered(D_{us_{k-2}} = \{x_i^{us_{k-2}}\}_{i=1}^n), x_i^{sk-1} \in R_{N_{sk-1}}$

for epoch = 1,2,3,..., k do

Update θ_c using Equation (20)–(22)

end for

Output: Predicted $\theta_c * Misclustered(D_{us_{k-2}} = \{x_i^{us_{k-2}}\}_{i=1}^n)$

Step 5: Classifier for fault diagnosis

In the DSVMN, the fault of mechanical equipment with $j = m$ classes is diagnosed using the extracted feature $\theta_c * Misclustered(D_{us_{k-2}} = \{x_i^{us_{k-2}}\}_{i=1}^n)$,

Clustered($D_{us_{k-2}} = \{x_i^{us_{k-2}}\}_{i=1}^n$)

for epoch = 1,2,3,..., k do

Update θ_s using Equations (23)–(25)

end for

Output: Predicted $\{\{y_i^{sk-2}\}_{i=1}^n\}'$

In this method, it is possible to obtain data in which weights have been added to feature points using the fault clustering results of the encoder output. The proposed method can solve data sparsity and data imbalance problems by correcting the feature values for mechanical equipment that are not extracted well in unsupervised learning using DNN models. It is possible to achieve generalization and robustness in industrial environments.

The weighted encoder output data not included in the target centroid and the encoder output data of the target centroid, are combined and input to the DSVMN classifier. A radial basis function (RBF) kernel-based SVM, which can classify multiple nonlinear feature spaces, has been widely used for detecting anomalies in mechanical equipment (Han, Zhang, Yin, & Tan, 2021; Miao, Zhang, Lin, Zhao, Liu, Liu, & Li, 2022; Wang, Yao, Chen, & Ding, 2021; Yao, Fang, Xiao, Hou, & Fu, 2021). The proposed DSVMN converts the existing SVM into a neural network. The input layer of this model is an RBF kernel layer that can

transform a nonlinear feature space into a linear classification space and maps the mechanical equipment feature data extracted from unsupervised learning to a Gaussian space. The feature values obtained through these layers pass through the second layer, a fully connected layer, the third layer, a dropout layer to prevent overfitting, and the last layer, with is a fully connected layer that classifies normal data and abnormal data. In the DSVMN model, normal data and abnormal data are classified according to the labels in the data using the features learned by the ensemble feature extractor. In addition, K-fold cross-validation is performed to prevent overfitting of the DSVMN model. In this way, a supervised learning method based on unsupervised data can solve the problem of domain adaptability for mechanical equipment data.

3.3. Feature extractor for unsupervised data

The proposed 2D CNN-LSTM-based AE model is trained on data $\{x_i^{us_{k-1}}\}_{i=1}^n$ to extract features in order to obtain a low-dimensional, high-level data space. $\{x_i^{us_{k-1}}\}_{i=1}^n$ comprises normal and abnormal data for the fault diagnosis of mechanical equipment. In an AE, features are extracted from the input using an encoder network. The compressed feature vector is restored to the size of the input vector through the decoder and a high-dimensional value is output. The input value is defined as $\{x_i^{us_{k-1}}\}_{i=1}^n$, the feature value extracted from the encoder is $z(\{x_i^{us_{k-1}}\}_{i=1}^n)$ value in Eq. (10), the encoder parameter is θ_e , the decoder output value x' can be expressed as Eq. (11) and the decoder parameter is θ_{de} . The decoder output x' is the output of the AE. The AE can be used to solve data sparsity problems by extracting unsupervised data features that reflect the underlying structure of the data. In addition, it can improve the adaptability of the model to mechanical equipment data with various characteristics.

$$z(\{x_i^{us_{k-1}}\}_{i=1}^n) = \{x_i^{us_{k-2}}\}_{i=1}^n \quad (10)$$

$$x' = h(z(\{x_i^{us_{k-1}}\}_{i=1}^n)) \quad (11)$$

In this study, the auto encoder extracts features of the time series-based mechanical equipment datasets using 2D CNNs and LSTMs. In the encoder, there are convolutional layers with 128, 64, and 32 filters, maxpooling layers, and an LSTM layer. The decoder has a structure symmetrical to that of the encoder, where an upsampling layer is used instead of the maxpooling layer. We optimize the loss function of AE L_{AE} to obtain the output, which is sent to the K-means, DNN, and DSVMN in the fully connected layer, where the bottleneck between the encoder and decoder occurs. In Eq. (12) and Eq. (13) for optimizing the auto-encoder loss function L_{AE} and parameters θ_e , and θ_{de} are as follows.

$$L_{AE} = \frac{1}{n} \sum_{i=1}^n (x_i^{us1} - x_i')^2 \quad (12)$$

$$\theta'_e, \theta'_{de} = \underset{\theta'_e, \theta'_{de}}{\operatorname{argmin}} L_{AE}(\theta_e, \theta_{de}) \quad (13)$$

Adadelta gradient descent was used to learn the unsupervised data by optimizing the parameters of the AE feature extractor.

$$G(t) \leftarrow \gamma G(t-1) + (1-\gamma) \left(\frac{\partial L_{AE}}{\partial (\theta_e, \theta_{de})} \right)^2 \quad (14)$$

$$\Delta(\theta_e, \theta_{de}) \leftarrow \frac{\sqrt{\Delta S(t-1) + \epsilon}}{\sqrt{G(t) + \epsilon}} * \frac{\partial L_{AE}}{\partial (\theta_e, \theta_{de})} \quad (15)$$

$$S(t) \leftarrow \gamma S(t-1) + (1-\gamma) \Delta(\theta_e, \theta_{de})^2 \quad (16)$$

$$\theta_e, \theta_{de} \leftarrow \theta_e, \theta_{de} - \Delta(\theta_e, \theta_{de}) \quad (17)$$

Here, $G(t)$ is the exponential mean function and γ represents the step size in Eq. (14). To adjust the learning rate of Adadelta gradient descent, we use $\Delta(\theta_e, \theta_{de})$ and $S(t)$ in Eq. (15) and Eq. (16), which decrease the

learning rate according to the change in weight Eq. (17). Adadelta can be used to obtain high-level feature vectors $\mathbf{z}(\{\mathbf{x}_i^{us_{k-1}}\}_{i=1}^n)$.

3.4. Feature validator for unsupervised data

We propose a K-means clustering method to validate the features of the unsupervised data. Distance measurement learning using the K-means clustering approach is used to validate the feature extraction. In this study, the feature extracted by the encoder is used as the input of the K-means model. This method contributes to the domain adaptability by evaluating sparse data. We use the low-dimensional, high-quality dataset $D_{us_{k-2}}$ obtained from the encoder to learn the Euclidean distance between the data and cluster centroids in Eq. (18). The cluster centroids are expressed as $\{\mu_i\}_{i=1}^{N_c}$ clusters, where N_c is the number clusters. In addition, the N_c of the high-quality feature vector is equal to the number of classes in the dataset. In this step, for verification, an iterative optimization method is proposed to evaluate the quality of the data by obtaining $Misclustered(D_{us_{k-2}} = \{\mathbf{x}_i^{us_{k-2}}\}_{i=1}^n)$, which are the data that do not belong to the target centroid according to the unsupervised data points, and $Clustered(D_{us_{k-2}} = \{\mathbf{x}_i^{us_{k-2}}\}_{i=1}^n)$, which have been clustered into the target centroid in Eq. (19). Unsupervised data with poor clustering, i.e., $Misclustered(D_{us_{k-2}} = \{\mathbf{x}_i^{us_{k-2}}\}_{i=1}^n)$, is used as the input value for centroid based learning. The clustering results on unsupervised data performed well because the features were extracted from the encoder using the mel spectrogram images. In addition, we initialized the k-means to cluster when the number of data samples allocated to the centroid of the k-means is less than 50 % of the number of samples corresponding to the class to ensure the stability of the model.

$$ED_{(i,j)} = \sum_{j=1}^{N_c} \sum_{i=1}^n \|\mathbf{x}_i^{us_{k-2}} - \mu_j\|^2 \quad (18)$$

$$D_{s_{k-1}} = \{\mathbf{x}_i^{s_{k-1}}\}_{i=1}^n \leftarrow Misclustered(D_{us_{k-2}} = \{\mathbf{x}_i^{us_{k-2}}\}_{i=1}^n), Clustered(D_{us_{k-2}} = \{\mathbf{x}_i^{us_{k-2}}\}_{i=1}^n) \quad (19)$$

3.5. Centroid based learning using unsupervised data

The $Misclustered(D_{us_{k-2}} = \{\mathbf{x}_i^{us_{k-2}}\}_{i=1}^n)$ data points that are not clustered into the target centroid were obtained through the K-means algorithm. These data points have unclear characteristics and can adversely affect the supervised learning-based fault diagnosis performance. Therefore, we propose a DNN algorithm that uses the Euclidean distance loss function to map the misclustered data points to the coordinates of the target centroid through weights. The proposed DNN consists of five fully connected layers, and each layer consists of 128, 64, 32, 16, and 300 nodes. The Eq. (20), Eq. (21), and Eq. (22) for the loss function and weight parameter θ_C for mapping the data point to the coordinates of the optimized target centroid are as follows.

$$LED_{(i,j)} = \sum_{j=1}^{N_c} \sum_{i=1}^n \|\mathbf{x}_i^{us_{k-1}} - \mu_j\|^2 \quad (20)$$

$$\theta'_C = argmin_{\theta_C} L_{AE}(\theta_C) \quad (21)$$

$$\theta_C \leftarrow \theta_C - \delta \frac{\partial L_{ED_{(i,j)}}}{\partial \theta_C} \quad (22)$$

3.6. Classification for fault diagnosis using unsupervised data

Classification is performed by the proposed supervised learning-based model using the misclustered dataset obtained from the centroid based learning DNN, the encoder's output values, and the clustered values. In the DSVMN model, mechanical equipment data are mapped in Gaussian space through the RBF kernel-based layer. It has the advantage of improving the accuracy of the mechanical equipment fault diagnosis classification with the features previously extracted through spatial transformation. Parameter γ of the RBF kernel is responsible for regulating the hyperparameters. In Eq. (23), γ as the nearest neighbor kernel, can adjust the distance between $\mathbf{x}_i^{s_{k-2}}$ and $\mathbf{x}_j^{s_{k-2}}$. In this experiment, γ is set to 0.001. The RBF kernel equation is as follows.

$$RBF = exp(-\gamma \|\mathbf{x}_i^{s_{k-2}} - \mathbf{x}_j^{s_{k-2}}\|^2) \quad (23)$$

In Eq. (24), s_j is the score of the incorrect mechanical equipment label, $s_{y_i^{s_{k-2}}}$ is the score of the true mechanical equipment label, and 1 is the value of the hinge loss function L_{svm} . L_{svm} is used to regularize soft margin SVM for fault diagnosis. The Eq. (25) of parameter θ_c uses the hinge loss L_{svm} for optimized DSVMN mechanical equipment fault diagnosis classification, and it is expressed follows.

$$L_{svm} = \sum_{j \neq y_i} max(0, s_j - s_{y_i^{s_{k-2}}} + 1) \quad (24)$$

$$\theta_c \leftarrow \theta_c - \delta \frac{\partial L_{svm}}{\partial \theta_c} \quad (25)$$

The number of nodes of the last fully connected layer is equal to the number of classes, and the classes in the multi-class mechanical equipment dataset are the predicted $\{\{\{\mathbf{y}_i^{s_{k-2}}\}_{i=1}^n\}_{j=1}^{m'}\}$.

4. Experimental study

In this study, data augmentation was performed on the mechanical equipment dataset to overcome the problems of data imbalance and data sparsity. In addition, experiments were performed using mel spectrogram images, which yield better performance than the raw vibration dataset. To consider the various environments of real industrial sites in the input values of the proposed ADTC model, five experiments were conducted. The proposed ADTC model achieved the best performance in various environments, proving its robustness.

4.1. Data description and algorithm setup

This section describes the CWRU bearing dataset, the MIMII dataset, and the Paderborn University bearing dataset used in the proposed method for the fault diagnosis of mechanical equipment.

4.1.1. CWRU bearing dataset

The data in the CWRU bearing dataset were collected using an accelerometer on faults in the drive and fan ends of the motor bearing: defects were present in the rolling elements, inner race, and outer race of the bearings. In this dataset, each load was configured using 0, 1, 2, or 3 hp at sampling frequencies of 12 or 48 kHz. In this experiment, only the high frequency 48 kHz sampling data were used to conduct the effect-

defect diagnosis study. Furthermore, the 1, 2, and 3 hp datasets were used in the experiment because there were missing values in the 0 hp dataset.

4.1.2. MIMII dataset

The MIMII dataset consists of industrial machine data for valves, pumps, fans, and slide rails. This dataset contains both normal and abnormal data with a time period of 10 s. The abnormal data contain contamination leakage, rotational imbalance, and rail damage information. All the sounds were recorded at 6, 0, or -6 dB noise at a 16 kHz sampling rate using a 16-bit microphone.

4.1.3. Paderborn University bearing dataset

The Paderborn University bearing dataset consists of normal data, data for inner race faults, and data for outer race faults. This dataset consists of 1500 RPM and 900 RPM data (Kimotho, Lessmeier, Sextro, & Zimmer, 2016). Table 1 shows the abnormal data for the outer race fatigue, which includes pitting and plastic deform indentations, and the inner race fatigue, which includes pitting. In addition, the data for the normal state, as listed in Table 2, presents various run-in period, radial load, and speed values. All vibration datasets were recorded at a 64 kHz sampling rate.

4.2. Data preprocessing

This section describes the sliding window augmentation method, audio data augmentation method, and SSIM image data augmentation verification method for overcoming the data sparsity of mechanical equipment datasets. In addition, the mel spectrogram images used in the fault diagnosis of mechanical equipment are explained.

Table 1
Paderborn University bearing normal data.

Bearing Code	Run-in Period	Radial Load [N]	Speed [min]	No. Samples
K001	>50	1000–3000	1500–2000	3969
K002	19	3000	2900	3969
K003	1	3000	3000	3969

Table 2
Paderborn University bearing abnormal data.

Bearing code	Damage (main mode and symptom)	Element type	Combination of damage	Arrangement	Extent of damage	Characteristic of damage	No. Samples
KA04	Fatigue: pitting	Outer Race	Single	no repetition	1	Single	3969
KA15	Plastic deform Indentations	Outer Race	Single	no repetition	1	Single	3969
KA16	Fatigue: pitting	Outer Race	Repetitive	random	2	Single	3969
KI04	Fatigue: pitting	Inner Race	Multiple	no repetition	1	Single	3969
KI14	Fatigue: pitting	Inner Race	Multiple	no repetition	1	Single	3969
KI16	Fatigue: pitting	Inner Race	Single	no repetition	3	Single	3969
KI18	Fatigue: pitting	Inner Race	Single	no repetition	2	Single	3969

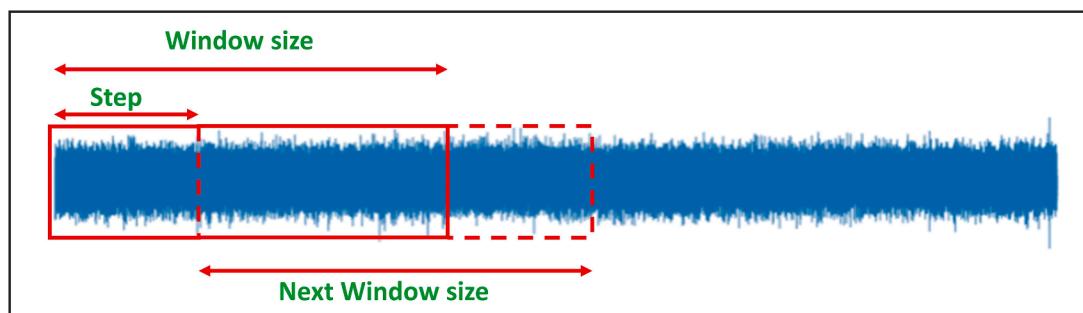


Fig. 2. Sliding window method for data augmentation.

industrial machine vibration dataset can be used in various experiments to creating datasets reflecting various environments. In the experiment, the augmented data and original data were used as the training dataset, and the experiment was conducted using a part of the original data as the test set.

4.2.2. Mel spectrogram

The time-frequency-based mel spectrogram image method is widely used in the field of data-driven fault diagnosis in the case of noisy datasets because models using mel spectrogram images have shown better performances as compared to the case in which raw signals were used (Hong & Suh, 2021; Hossain, 2019). Therefore, the augmented raw signals of the bearing dataset and the industrial machine dataset were converted into a mel spectrogram image in this experiment. The hyperparameters used for this conversion were the sampling rate, mel band, frame length, and frame stride. Mel bands of 80 and 40 were used for the bearing dataset with a 48 kHz sampling rate and an industrial machine dataset with a 16 kHz sampling rate. Additionally, frame lengths and frame strides of 0.025 and 0.010 s, respectively, were used. The raw signal and converted mel spectrogram images are depicted in Fig. 3. In the mel spectrogram images, the x-axis indicates time and the y-axis indicates the frequency. To express the raw signal as a mel spectrogram image, a fast Fourier transform is performed, and the window segment is mapped from the time to the frequency domain through the parameters. The amplitude of the raw signal is converted into decibels and mapped to the mel scale. The y-axis of the mel spectrogram represents the mel domain from 0 to 8192, and the x-axis represents the time of the data to express the characteristics of the raw signal data as an image. By representing it as an image, it is possible to extract easily recognizable features of data via machine learning.

4.2.3. SSIM

The structural similarity index method (SSIM) is used to measure the similarity to the original image with respect to the distortions caused by compression and transformation. This method was used in this study to compare the industrial machine image data to which audio augmentation was applied to the original images. The Eq. (26) used in this method can be expressed as follows (Channappayya, Bovik, & Heath, 2008):

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(2\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (26)$$

where μ_x , μ_y , σ_x^2 , σ_y^2 , and σ_{xy} represent the mean, variance, and covariance of x and y , respectively, and c_1 and c_2 are the normalization constant and the contrast term. Based on the comparison of the original and augmented image data using the SSIM method, only the dataset with an SSIM of 90 % or higher was used as the input value of the DNN model. Table 4 lists the number of augmented industrial machine dataset samples.

4.3. Numerical experiment

In order to evaluate the ADTC model based on mel spectrogram image inputs, experiments were performed to evaluate the mel spectrogram images, domain adaptability, bearing and industrial machine datasets, bearing analysis for different rotational speeds, and damage analysis. Various noise levels, lengths, loads, and RPM values were applied to the mechanical equipment dataset in the experiments. The results show that the proposed ADTC model achieved the best performance in the various cases.

4.3.1. Performance analysis of mel spectrogram images

Many studies have revealed that the accuracy of deep learning can be improved through data feature extraction in the time-frequency domain by transforming the raw signals into mel spectrogram images (Oh, 2020; Pandey, Shekhawat, & Prasanna, 2019). Accordingly, a comparative

analysis according to data type was performed in this experiment using the ADTC model to establish that the fault diagnosis-based mel spectrogram images exhibited a better performance than those based on a fast Fourier transform (FFT) and raw signals. The raw signal, FFT, and mel spectrogram image datasets were obtained under the same conditions. The bearing dataset for the comparative analysis consisted of 1, 2, and 3 hp and the industrial machine dataset was used with -6, 0, and 6 dB noise. The ADTC model in the experiment was used as a 1D convolutional layer of AE according to the FFT and raw signal dimension and as a 2D layer for the mel spectrogram image. The accuracy according to data type is presented in Table 5. Evidently, the raw data had a mean accuracy of 81 %, whereas the mel spectrogram image data showed a mean accuracy of 98 %. The two types of datasets show similar trends corresponding to the respective conditions; the bearing data show similar results for loads and the performance accuracy of the industrial machine data generally tends to decrease when a high volume of noise data is present. Overall, the results establish that the mel spectrogram image data are superior to the raw data. Therefore, these experimental results prove the superiority and necessity of diagnosing mechanical equipment faults using the mel spectrogram image. Moreover, the proposed mel spectrogram image-based fault diagnosis method surpasses the existing raw signal and FFT-based method and shows high potential. Clearly, fault diagnosis using time and frequency-based mel spectrogram images of bearing and industrial machine datasets can improve model accuracy.

4.3.2. Domain adaptation performance analysis

In this experiment, the domain adaptation of the proposed mel spectrogram image-based ADTC model was evaluated with different bearing loads. A single load sample was used as shown in Table 1, and an experiment was performed to compare and analyze domain adaptation tests of other models. Table 6 shows the performance of the proposed ADTC model, SVM (Konar & Chattopadhyay, 2011), CNN-SVM (Xu, Ma, Zhang, Yang, Li, & Liu, 2019), denoising AE (DAE) (Li et al., 2019), deep representation clustering (DRC) (Li et al., 2020), and the models based on the results of the comparative analysis. The models, including the SVM, exhibited a good overall performance. Additionally, each model exhibited the robust adaptability to other loads when a 2 hp load is learned. When transfer learning was performed with each other, good results were obtained, whereas those for 1 and 3 hp showed poor results. In the prediction of each class in transfer learning, the remaining classes showed excellent predictions except for Ball 0.014 Fault and Ball 0.021 Fault. These results indicate that the Ball 0.014 Fault and Ball 0.021 Fault classes can become a factor in the performance degradation of fault diagnosis when predicted through transfer learning using different loads. Compared with the other models, the proposed ADTC model exhibited the best results when transfer learning for each load was performed. Fig. 4 shows the confusion matrices for the transfer learning of the proposed model. The confusion matrix shows the prediction accuracy of the model by comparing the predicted label and the actual label for transfer learning according to the load of each bearing. It shows the best performance of the transfer learning experiment from 2 hp to 1 hp, and shows the degraded performance of the model from 3 hp to 1 hp. As mentioned above, the predictions of Ball 0.014 Fault and Ball 0.021 Fault affect the accuracy.

4.3.3. Bearing dataset analysis considering different noise levels

Data obtained from real-life industrial locations using sensors are generally contaminated with noise, which reduces the accuracy of the data-driven fault diagnosis model (Tagawa, Maskeliūnas, & Damasevičius, 2021). Hence, a robustness test of the proposed mel spectrogram image-based model was conducted for various noise environments. The bearing dataset for the robustness test was constructed by applying noise to the original data corresponding to the following signal-to-noise ratios (SNRs): -8, -6, -4, -2, 0, 2, 4, 6, and 8 dB. The Eq. (27) for the SNR signal is as follows:

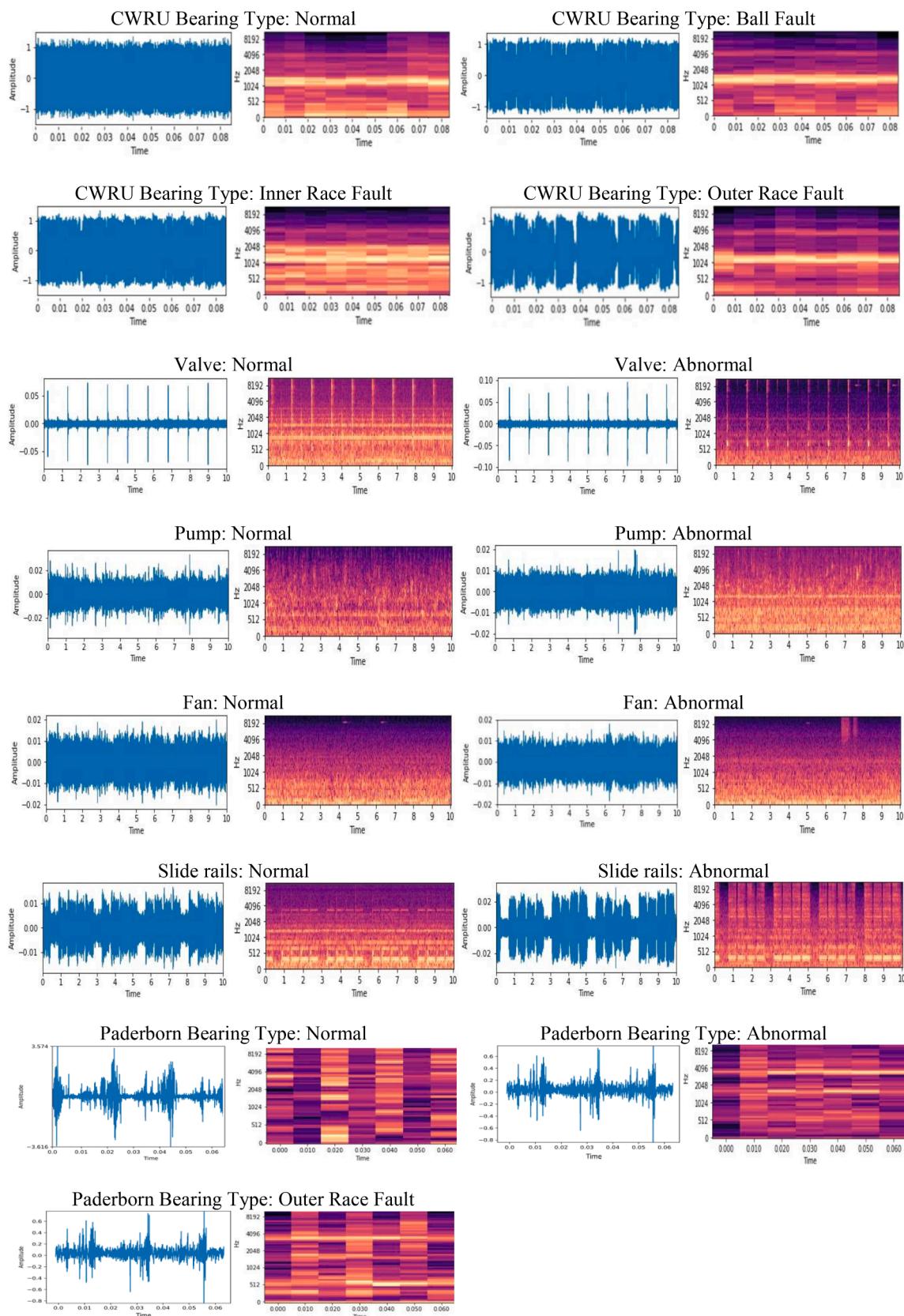
**Fig. 3.** Original raw signals and mel spectrogram images.

Table 4
MIMII dataset.

Data type	SNR (dB)	Fault type	No. of datasets	No. of augmented datasets
Valve	6	Normal	3691	—
	6	Abnormal	479	2395
	0	Normal	3691	—
	0	Abnormal	479	2395
	-6	Normal	2699	—
	-6	Abnormal	479	2395
	6	Normal	3749	—
	6	Abnormal	456	2280
Pump	0	Normal	3749	—
	0	Abnormal	456	2280
	-6	Normal	3749	—
	-6	Abnormal	456	2280
Fan	6	Normal	4075	—
	6	Abnormal	1475	4075
	0	Normal	4075	—
	0	Abnormal	1475	4075
Slide Rails	-6	Normal	4075	—
	-6	Abnormal	1475	4075
	6	Normal	3204	—
	6	Abnormal	890	3204
	0	Normal	3204	—
	0	Abnormal	890	3204
	-6	Normal	3204	—
	-6	Abnormal	890	3204

$$SNR_{dB} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \quad (27)$$

where P_{signal} represents the average signal power and P_{noise} represents the average noise power. We added white Gaussian noise according to the SNR for each dataset to increase the noise levels of all of the samples for each load. The performance evaluation of the proposed and comparative models is shown in Table 7. Generally, the accuracy of the model decreased as the noise-power ratio increased to -8 dB. Conversely, as the signal-power ratio became higher than the noise-power ratio and

Table 5
Accuracy (%) obtained using raw signals, FFT, and mel spectrogram images.

Data type	Pump		
SNR	-6 dB	0 dB	6 dB
Signal	81	82	86
FFT	97	98	98
Image	98	97	99
Data type	Fan		
SNR	-6 dB	0 dB	6 dB
Signal	73	79	83
FFT	98	99	99
Image	100	99	100
Data type	Slide Rails		
SNR	-6 dB	0 dB	6 dB
Signal	68	90	87
FFT	98	99	99
Image	99	99	99
Data type	Valve		
SNR	-6 dB	0 dB	6 dB
Signal	82	85	80
FFT	96	98	98
Image	99	99	99
Data type	Bearing		
Load	1 hp	2 hp	3 hp
Signal	89	89	85
FFT	98	98	99
Image	99	99	98

Table 6
Model accuracy (%) results obtained using different loads.

Method	1 hp to 2 hp	1 hp to 3 hp	2 hp to 1 hp	2 hp to 3 hp	3 hp to 1 hp	3 hp to 2 hp
SVM	94	76	95	93	83	96
CNN-SVM	95	66	95	94	72	83
DAE	91	72	72	75	73	71
DRC	84	76	80	83	65	80
Proposed Model	96	97	98	97	95	97

approached 8 dB, the accuracy of the model tended to increase. In conclusion, bearing data containing noise can interfere with the diagnosis of inner and outer race defects of the bearing. However, through comparison with other models, it was proven that the proposed ADTC model is robust and less susceptible to noise in real-life industrial environments.

4.3.4. Industrial machine dataset analysis considering different noise

An experiment was performed to consider the effect of noise in a time-series dataset of industrial machines. A large number of data samples tested previously produced significantly reduced accuracy of the model because of the imbalance in the data. Therefore, in this study, the fault diagnosis of an industrial machine was performed using the augmented dataset presented in Table 4. The proposed mel spectrogram image-based model was evaluated using precision, recall, and F1 scores while taking four comparative models and two data augmentation methods into consideration: the corresponding results are listed in Table 8. Two data augmentation methods, as presented in Table 4, were applied: Case 1, where only the abnormal data were augmented to resolve the data imbalance, and Case 2, where both the normal and abnormal data were augmented, but the data imbalance could not be resolved. Both normal and abnormal data construct the number of samples, for which the data were augmented by applying five multiples of the original number of samples. Each of the models in which the noise-power ratio was increased showed poor results, similar to that in the case of bearing. Depending on the data augmentation method, the two methods show similar results, or the data augmentation method for solving the imbalance problem showed a higher classification performance. Therefore, for the industrial machine dataset, the balanced data augmentation method is recommended for both normal and abnormal datasets because the balanced dataset exhibits better performance. Overall, the proposed model as well as the supervised learning-based SVM model produced desirable results. Additionally, the industrial machine dataset exhibited differences based on type; the smallest difference was observed in the valve dataset, whereas the fan data were found to be vulnerable to noise. In conclusion, the proposed ADTC model utilizing the data preprocessing method proved to be the most robust for noise reduction and exhibited the highest accuracy. Fig. 5 indicates the classification performance corresponding to Case 1 of the proposed model with the best results. The AUC value, i.e., the area under the ROC curve, corresponding to the ROC area signifies classification performance. The ROC curve shows the performance of the classification model at various thresholds through the metrics of true positive rate (TPR) and false positive rate (FPR). The range of AUC values is between 0 and 1, and when the prediction is 100 % correct, the AUC is 1.0, and an AUC of a model with 100 % incorrect predictions is 0. The proposed model has an AUC of 1.0, indicating excellent prediction accuracy for industrial machines.

4.3.5. Bearing dataset analysis considering different rotational speeds and damage

An experiment was performed to consider the effect of rotational speed and load on the time series bearing dataset. Existing bearing datasets are small in size. Therefore, in this study, bearing faults were diagnosed using an augmented dataset obtained using the sliding

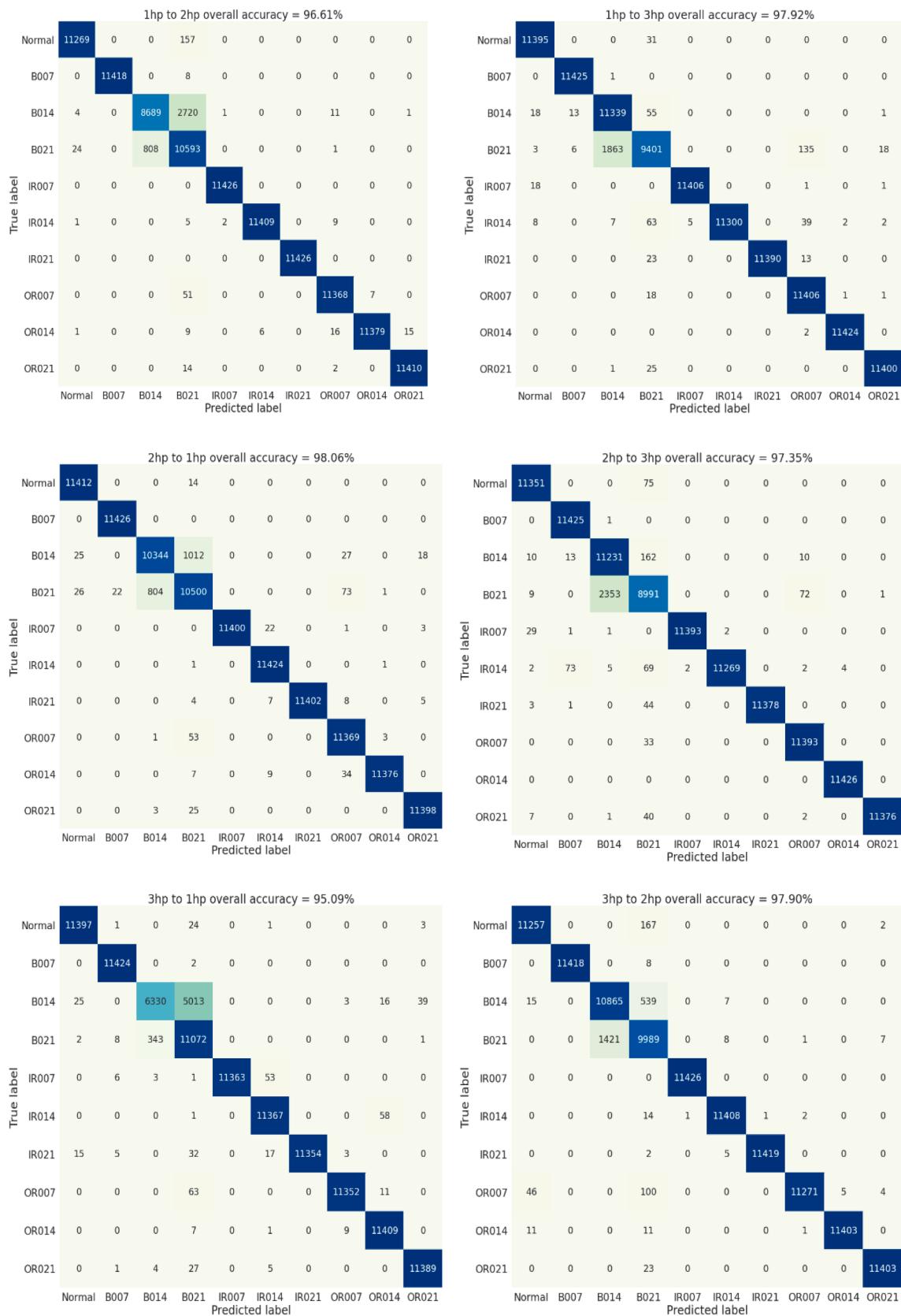


Fig. 4. Confusion matrices of different loads.

Table 7

Model accuracy (%) based on the noisy bearing dataset.

Dataset	1 hp Bearing								
SNR	-8 dB	-6 dB	-4 dB	-2 dB	0 dB	2 dB	4 dB	6 dB	8 dB
SVM	90	91	92	90	94	96	96	96	97
CNN	90	91	90	92	95	98	97	96	98
SVM									
DAE	75	72	76	75	78	78	92	90	90
DRC	72	73	73	72	74	83	82	86	87
ADTC Model	97	97	97	96	99	99	99	99	99
Dataset	2 hp Bearing								
SNR	-8 dB	-6 dB	-4 dB	-2 dB	0 dB	2 dB	4 dB	6 dB	8 dB
SVM	92	91	92	90	94	96	96	96	96
CNN	88	93	93	93	95	97	98	98	98
SVM									
DAE	75	73	74	75	90	90	90	91	92
DRC	72	74	74	72	72	74	80	83	84
ADTC Model	96	95	96	97	98	99	99	99	99
Dataset	3 hp Bearing								
SNR	-8 dB	-6 dB	-4 dB	-2 dB	0 dB	2 dB	4 dB	6 dB	8 dB
SVM	86	92	92	90	94	96	96	96	96
CNN	93	94	93	92	93	92	96	98	98
SVM									
DAE	72	73	73	80	80	83	82	84	90
DRC	72	73	73	80	80	80	80	83	84
ADTC Model	98	97	98	99	99	99	99	99	99

Table 8

Model accuracy (%) based on noisy the industrial machine dataset.

Dataset	Valve dataset								
SNR	-6 dB			0 dB			6 dB		
Method	P	R	F1	P	R	F1	P	R	F1
SVM	0.92	0.92	0.92	0.90	0.90	0.90	0.96	0.96	0.96
CNN-SVM	0.91	0.92	0.91	0.92	0.92	0.92	0.99	0.99	0.99
DAE	0.86	0.86	0.86	0.85	0.85	0.85	0.92	0.94	0.92
DRC	0.89	0.89	0.89	0.88	0.89	0.88	0.92	0.92	0.92
ADTC Model	Case 1	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	Case 2	0.95	0.95	0.95	0.95	0.94	0.95	0.96	0.96
Dataset	Fan dataset								
SNR	-6 dB			0 dB			6 dB		
Method	P	R	F1	P	R	F1	P	R	F1
SVM	0.90	0.90	0.90	0.94	0.94	0.94	0.98	0.98	0.98
CNN-SVM	0.94	0.94	0.94	0.93	0.94	0.93	0.98	0.98	0.98
DAE	0.85	0.86	0.85	0.83	0.85	0.83	0.84	0.90	0.86
DRC	0.88	0.88	0.88	0.88	0.88	0.88	0.96	0.96	0.96
ADTC Model	Case 1	1.00	1.00	1.00	0.99	0.98	0.98	1.00	0.99
	Case 2	0.95	0.95	0.95	0.96	0.95	0.95	0.96	0.95
Dataset	Slide rails dataset								
SNR	-6 dB			0 dB			6 dB		
Method	P	R	F1	P	R	F1	P	R	F1
SVM	0.86	0.86	0.86	0.88	0.88	0.94	0.95	0.94	0.94
CNN-SVM	0.83	0.84	0.83	0.91	0.91	0.91	0.90	0.92	0.90
DAE	0.83	0.85	0.83	0.80	0.84	0.81	0.89	0.92	0.90
DRC	0.93	0.93	0.93	0.92	0.92	0.92	0.93	0.94	0.93
ADTC Model	Case 1	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99
	Case 2	0.96	0.96	0.96	0.95	0.95	0.98	0.98	0.98
Dataset	Pump dataset								
SNR	-6 dB			0 dB			6 dB		
Method	P	R	F1	P	R	F1	P	R	F1
SVM	0.84	0.86	0.85	0.93	0.93	0.93	0.95	0.95	0.95
CNN-SVM	0.82	0.84	0.82	0.94	0.94	0.94	0.97	0.97	0.97
DAE	0.82	0.83	0.82	0.84	0.86	0.84	0.88	0.88	0.88
DRC	0.86	0.86	0.86	0.90	0.90	0.90	0.93	0.92	0.92
ADTC Model	Case 1	0.98	0.98	0.98	0.98	0.97	0.97	0.99	0.99
	Case 2	0.96	0.96	0.96	0.98	0.99	0.98	0.99	0.99

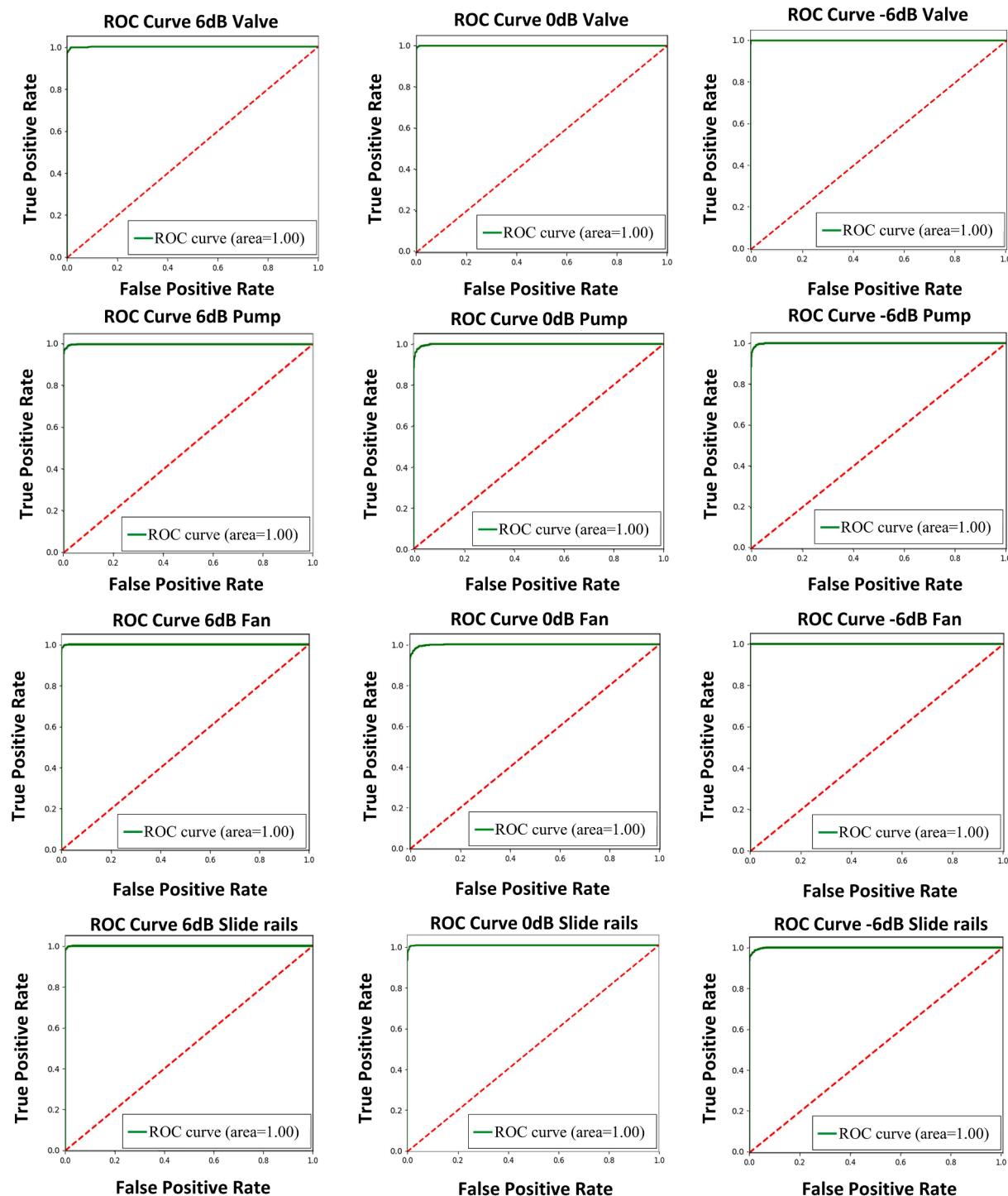


Fig. 5. ROC curve of the proposed model based on the industrial machine dataset.

Table 9
Model performance for the RPM of bearing dataset.

Dataset	Bearing dataset											
	1500 RPM to 900 RPM			900 RPM to 1500 RPM			1500 RPM			900 RPM		
Method	P	R	F1	P	R	F1	P	R	F1	P	R	F1
SVM	0.99	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.99
CNN-SVM	0.95	0.96	0.95	0.95	0.95	0.96	0.95	0.95	0.95	1.00	1.00	1.00
DAE	0.93	0.94	0.93	0.94	0.95	0.95	0.95	0.95	0.95	0.98	0.98	0.98
DRC	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.98
ADTC model	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

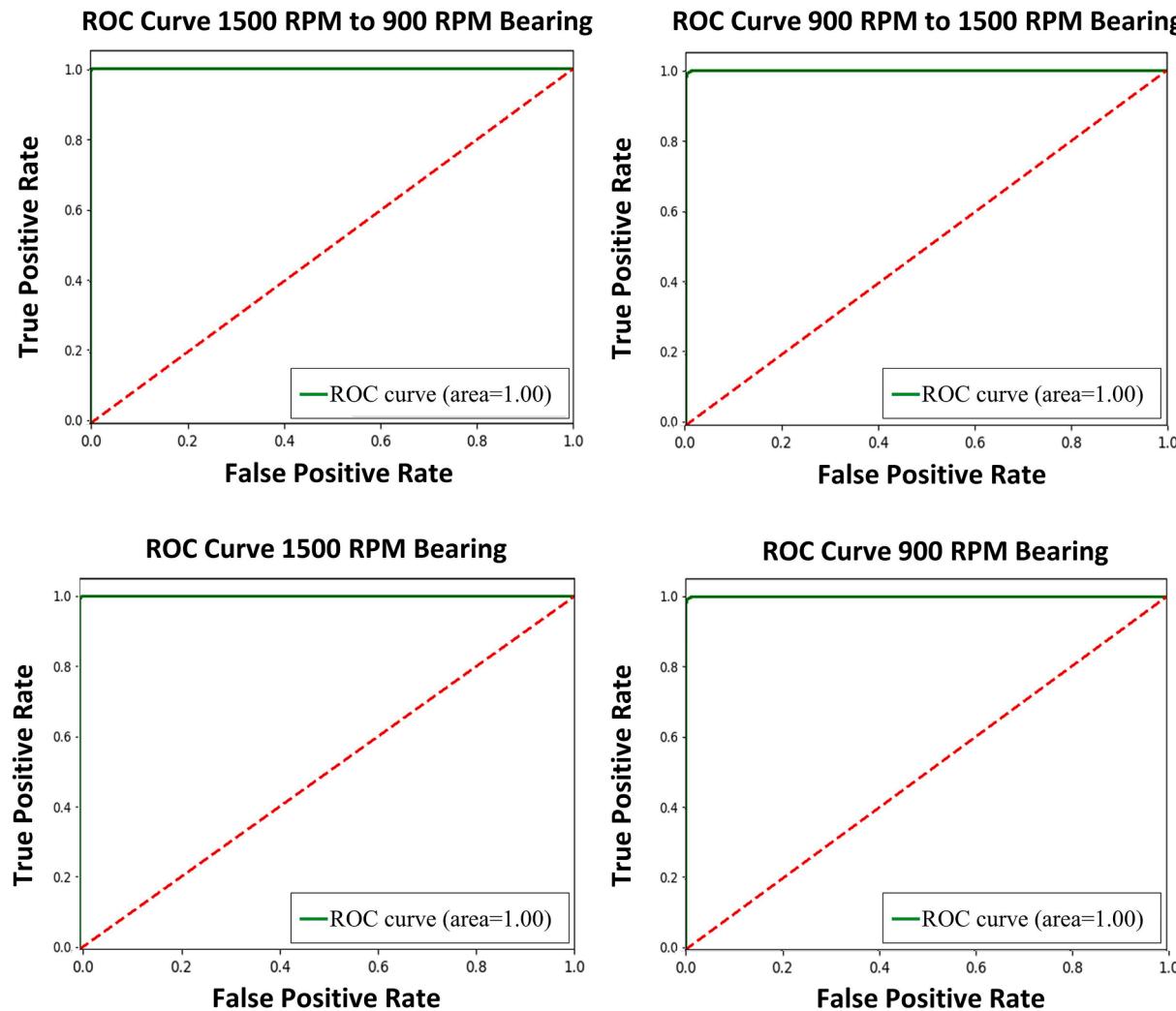


Fig. 6. ROC curve of the proposed model based on the bearing dataset.

window technique. These datasets are described in Table 1 and Table 2. The proposed mel spectrogram image-based model was evaluated using the four comparative metrics of precision, recall, and F1 score. The results are presented in Table 9. Experiments were carried out under the following four conditions: 1500 RPM to 900 RPM, 900 RPM to 1500 RPM, 900 RPM, and 1500 RPM. In addition, the dataset includes three sets of normal data with different run-in periods, radial loads, and speeds; IR(Inner Race) defect data with different extents of damage and damage combinations; and OR (Outer Race) defect data with different extents of damage, damage types, and damage combinations, yielding a total of 10 classes. The experiment was performed using this 10-class dataset. The results of the experiment reveal that the F1 score of the proposed ADTC model is the highest. In addition, the ROC curve of the ADTC model in Fig. 6 shows that it is possible to solve the problem of domain adaptability for various bearing conditions by obtaining a high AUC. Through experiments considering the rotational speed and damage of these bearings, the ADTC model is shown to have robustness, even on datasets with various types of normal state. Moreover, it shows its superiority in domain adaptability to other bearing types. Overall, the transfer learning for different RPMs and model predictions for a single RPM are consistent with the real labels. In addition, the proposed model shows high performance, even though there are several normal states.

4.3.6. Discussion

In this case study, five major cases were considered in the

experiments. In the first case, the performance of the proposed model was compared and analyzed using a time-series-based raw signal dataset for mechanical equipment, an FFT dataset, and a time-frequency-based mel spectrogram dataset, which were obtained by converting the raw signal. The analysis results indicate that the proposed mel spectrogram image-based fault diagnosis surpasses the existing raw signal and FFT methods and therefore shows promise. Clearly, fault diagnosis using time/frequency-based mel spectrogram images for bearing and industrial machine datasets can improve model accuracy. Moreover, because the model trained on the augmented dataset performs better, a data augmentation method for fault diagnosis is recommended.

In the second case, the domain adaptation of the proposed model was evaluated through a domain adaptation test for each load of the bearing. The best results were obtained when the bearing loads at 2 hp were learned. Experimental results indicate that the Ball 0.014 Fault and Ball 0.021 Fault classes can degrade fault diagnosis when predictions are made through transfer learning using different loads. In addition, the experimental results show that the proposed model has adaptability to several domains.

In the third case, we conducted a study on the robustness of the proposed model in a real industrial environment using the bearing dataset. As the noise-to-power ratio increased to -8 dB, the accuracy of the model decreased, and a comparison with other models revealed that the proposed ADTC model is robust and less sensitive to noise in real industrial environments.

In the fourth case, the robustness of the proposed model was evaluated by considering a real industrial environment using the industrial machine dataset. It was observed that the fault diagnosis performance of the mechanical equipment dataset is affected by the noise levels and the model based on an SVM exhibited a higher accuracy. For the industrial machine dataset, a balanced data augmentation method is recommended for both the normal and abnormal datasets because the balanced dataset leads to better performance.

In the fifth case, bearing data consisting of 10 classes (e.g., normal data with different run-in periods, radial loads, and speeds; IR defect data with different damage combinations, extents of damage, types of damage; and OR defect data with different damage combinations) were used in the experiments. Through experiments on rotational speed and bearing damage, it was shown that the ADTC model is robust even in datasets with various types of normal state and shows superiority in bearing-type domain adaptability to other methods.

5. Conclusion

In this paper, a mel spectrogram-based ADTC model that can improve system productivity and reliability by diagnosing abnormalities in an industrial environment was proposed. The main contributions of the proposed method are as follows. Data augmentation was performed to overcome the problem associated with the imbalanced time-series dataset of mechanical equipment. The limited ability to extract features from mechanical equipment data with sparsity was addressed by exploring feature extraction using an unsupervised learning-based model. Moreover, fault diagnosis was performed by including the extracted feature values of the unsupervised data in the supervised learning model.

Using the proposed method, the domain adaptive performances of the mechanical equipment and robustness under the influence of industrial noise were evaluated. The results indicate that the high predictive accuracy of the proposed fault-diagnosis model makes it a robust model for noisy environments and exhibits the adaptive performance of domains in datasets with various time lengths. Thus, it was demonstrated that the mel spectrogram-based ADTC model is an effective fault diagnosis approach and the balanced dataset has better fault-diagnosis performance. However, the K-means method has a limitation in that the number of clusters must be defined in advance.

In future study, we will use K-means to automatically designate the number of clusters. Domain adaptation experiments considering various environmental conditions and equipment types can also be conducted, in addition to diagnosis prediction studies for overcoming high data sparsity.

CRediT authorship contribution statement

Geonkyo Hong: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization.
Dongjun Suh: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Dongjun Suh reports article publishing charges was provided by National Research Foundation of Korea. Dongjun Suh reports article publishing charges was provided by Korea Institute of Energy Technology Evaluation and Planning and the Ministry of Trade, Industry & Energy of the Republic of Korea.

The remaining authors declare that they have no known competing

financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2021R1A5A8033165), (No. NRF-2021R1I1A3049503); the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20224000000150).

References

- Arias-Vergara, T., Klumpp, P., Vasquez-Correa, J. C., Nöth, E., Orozco-Arroyave, J. R., & Schuster, M. (2021). Multi-channel spectrograms for speech processing applications using deep learning methods. *Pattern Analysis and Applications*, 24(2), 423–431. <https://doi.org/10.1007/s10044-020-0921-5>
- Brito, L. C., Susto, G. A., Brito, J. N., & Duarte, M. A. V. (2022). An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery. *Mechanical Systems and Signal Processing*, 163(June 2021), 108105. <https://doi.org/10.1016/j.ymssp.2021.108105>
- Chalapathy, R., & Chawla, S. (2019). *Deep Learning for Anomaly Detection: A Survey*. 1–50.
- Channappayya, S. S., Bovik, A. C., & Heath, R. W. (2008). Rate bounds on SSIM index of quantized images. *IEEE Transactions on Image Processing*, 17(9), 1624–1639. <https://doi.org/10.1109/TIP.2008.2001400>
- Chen, P., Zhao, R., He, T., Wei, K., & Yang, Q. (2022). Unsupervised domain adaptation of bearing fault diagnosis based on Join Sliced Wasserstein Distance. *ISA Transactions*, 129, 504–519. <https://doi.org/10.1016/j.isatra.2021.12.037>
- Chen, Y., & Jin, H. (2019). Rare sound event detection using deep learning and data augmentation. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2019-Septe, 619–623. <https://doi.org/10.21437/Interspeech.2019-1985>
- Fahim, M., & Sillitti, A. (2019). Anomaly detection, analysis and prediction techniques in IoT environment: A systematic literature review. *IEEE Access*, 7, 81664–81681. <https://doi.org/10.1109/ACCESS.2019.2921912>
- Han, T., Zhang, L., Yin, Z., & Tan, A. C. C. (2021). Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine. *Measurement: Journal of the International Measurement Confederation*, 177(February), Article 109022. <https://doi.org/10.1016/j.measurement.2021.109022>
- Hong, G., & Suh, D. (2021). Supervised-learning-based intelligent fault diagnosis for mechanical equipment. *IEEE Access*, 9, 116147–116162. <https://doi.org/10.1109/access.2021.3104189>
- Hossain, M. S., & Muhammad, G. (2019). Emotion recognition using deep learning approach from audio-visual emotional big data. *Information Fusion*, 49(August 2018), 69–78. <https://doi.org/10.1016/j.inffus.2018.09.008>
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0192-5>
- Kimotho, J. K., Lessmeier, C., Sextro, W., & Zimmer, D. (2016). Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark dataset for data-driven classification. *Third European Conference of the Prognostics and Health Management Society 2016, Cm*, 152–156. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1088.9087&rep=rep1&type=pdf>.
- Konar, P., & Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). *Applied Soft Computing Journal*, 11(6), 4203–4211. <https://doi.org/10.1016/j.asoc.2011.03.014>
- Li, F., Liu, M., Zhao, Y., Kong, L., Dong, L., Liu, X., & Hui, M. (2019). Feature extraction and classification of heart sound using 1D convolutional neural networks. *EURASIP Journal on Advances in Signal Processing*, 2019(1). <https://doi.org/10.1186/s13634-019-0651-3>
- Li, X., Li, J., Qu, Y., & He, D. (2020). Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning. *Chinese Journal of Aeronautics*, 33(2), 418–426. <https://doi.org/10.1016/j.cja.2019.04.018>
- Li, X., Li, X., & Ma, H. (2020). Deep representation clustering-based fault diagnosis method with unsupervised data applied to rotating machinery. *Mechanical Systems and Signal Processing*, 143. <https://doi.org/10.1016/j.ymssp.2020.106825>
- Li, X., Zhang, W., & Ding, Q. (2019). Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. *Signal Processing*, 161, 136–154. <https://doi.org/10.1016/j.sigpro.2019.03.019>
- Li, X., Zhang, W., Ding, Q., & Sun, J. Q. (2019). Multi-Layer domain adaptation method for rolling bearing fault diagnosis. *Signal Processing*, 157, 180–197. <https://doi.org/10.1016/j.sigpro.2018.12.005>
- Li, X., Hu, Y., Zheng, J., Li, M., & Ma, W. (2021). Central moment discrepancy based domain adaptation for intelligent bearing fault diagnosis. *Neurocomputing*, 429, 12–24. <https://doi.org/10.1016/j.neucom.2020.11.063>

- Liu, H., Zhou, J., Xu, Y., Zheng, Y., Peng, X., & Jiang, W. (2018). Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks. *Neurocomputing*, 315, 412–424. <https://doi.org/10.1016/j.neucom.2018.07.034>
- Liu, S. C., & Liu, S. Y. (2003). An efficient expert system for machine fault diagnosis. *International Journal of Advanced Manufacturing Technology*, 21(9), 691–698. <https://doi.org/10.1007/s00170-002-1389-9>
- Miao, Y., Zhang, B., Lin, J., Zhao, M., Liu, H., Liu, Z., & Li, H. (2022). A review on the application of blind deconvolution in machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 163(June 2021), 108202. <https://doi.org/10.1016/j.ymssp.2021.108202>
- Mikolajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. *2018 International Interdisciplinary PhD Workshop, IPhDW 2018*, 117–122. <https://doi.org/10.1109/IPHDW.2018.8388338>.
- Oh, W. (2020). Comparison of environmental sound classification performance of convolutional neural networks according to audio preprocessing methods. *Journal of the Acoustical Society of Korea*, 39(3), 143–149. <https://doi.org/10.7776/ASK.2020.39.3.143>
- Pandey, S. K., Shekharawat, H. S., & Prasanna, S. R. M. (2019). Deep learning techniques for speech emotion recognition: A review. In *2019 29Th International Conference Radioelektronika, RADIOELEKTRONIKA 2019 - Microwave and Radio Electronics Week*. <https://doi.org/10.1109/RADIOELEK.2019.8733432>
- Pang, G., Shen, C., Cao, L., & Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys*, 54(2). <https://doi.org/10.1145/3439950>
- Qian, Q., Qin, Y., Wang, Y., & Liu, F. (2021). A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis. *Measurement: Journal of the International Measurement Confederation*, 178(March), Article 109352. <https://doi.org/10.1016/j.measurement.2021.109352>
- Rajoub, B. (2020). Supervised and unsupervised learning. *Biomedical Signal Processing and Artificial Intelligence in Healthcare*, January, 51–89. <https://doi.org/10.1016/b978-0-12-18946-7.00003-2>.
- Rauber, T. W., da Silva Loca, A. L., Boldt, F. de A., Rodrigues, A. L., & Varejão, F. M. (2021). An experimental methodology to evaluate machine learning methods for fault diagnosis based on vibration signals. *Expert Systems with Applications*, 167 (September 2020). 10.1016/j.eswa.2020.114022.
- Saufi, S. R., Ahmad, Z. A. Bin, Leong, M. S., & Lim, M. H. (2020). Gearbox Fault Diagnosis Using a Deep Learning Model with Limited Data Sample. *IEEE Transactions on Industrial Informatics*, 16(10), 6263–6271. <https://doi.org/10.1109/TII.2020.2967822>.
- Shenfield, A., & Howarth, M. (2020). A novel deep learning model for the detection and identification of rolling element-bearing faults. *Sensors (Switzerland)*, 20(18), 1–24. <https://doi.org/10.3390/s20185112>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0197-0>
- Shyu, M., Chen, S., & Iyengar, S. S. (2020). A survey on deep learning techniques. *Strad Research*, 7(8). <https://doi.org/10.37896/sr7.8/037>.
- Sohail, M., Kim, C. H., & Kim, J. M. (2017). A hybrid feature model and deep-learning-based bearing fault diagnosis. *Sensors (Switzerland)*, 17(12). <https://doi.org/10.3390/s17122876>
- Stevenson, M., Mues, C., & Bravo, C. (2021). Deep residential representations: Using unsupervised learning to unlock elevation data for geo-demographic prediction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 187(November 2021), 378–392. <https://doi.org/10.1016/j.isprsjprs.2022.03.015>
- Sun, J., Yan, C., & Wen, J. (2018). Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Transactions on Instrumentation and Measurement*, 67(1), 185–195. <https://doi.org/10.1109/TIM.2017.2759418>
- Tagawa, Y., Maskeliūnas, R., & Damaševičius, R. (2021). Acoustic anomaly detection of mechanical failures in noisy real-life factory environments. *Electronics (Switzerland)*, 10(19). <https://doi.org/10.3390/electronics10192329>
- Tao, H., Wang, P., Chen, Y., Stojanovic, V., & Yang, H. (2020). An unsupervised fault diagnosis method for rolling bearing using STFT and generative neural networks. *Journal of the Franklin Institute*, 357(11), 7286–7307. <https://doi.org/10.1016/j.jfranklin.2020.04.024>
- Tavakoli, A., & Heydarian, A. (2022). Multimodal driver state modeling through unsupervised learning. *Accident Analysis and Prevention*, 170(February), Article 106640. <https://doi.org/10.1016/j.aap.2022.106640>
- Tran, T., & Lundgren, J. (2020). Drill fault diagnosis based on the scalogram and MEL spectrogram of sound signals using artificial intelligence. *IEEE Access*, 8, 203655–203666. <https://doi.org/10.1109/ACCESS.2020.3036769>
- Tung, T. V., & Yang, B.-S. (2009). Machine fault diagnosis and prognosis: The state of the art. *International Journal of Fluid Machinery and Systems*, 2(1), 61–71. <https://doi.org/10.5293/ijfms.2009.2.1.061>
- Varga, A. (2017). Fault diagnosis. *Studies in Systems, Decision and Control*, 84(3), 27–56. https://doi.org/10.1007/978-3-319-51559-5_3
- Wang, J., Xue, M., Culhane, R., Diao, E., Ding, J., & Tarokh, V. (2020). Speech emotion recognition with dual-sequence LSTM architecture. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6469–6473.
- Wang, Z., Yao, L., Chen, G., & Ding, J. (2021). Modified multiscale weighted permutation entropy and optimized support vector machine method for rolling bearing fault diagnosis with complex signals. *ISA Transactions*, 114, 470–484. <https://doi.org/10.1016/j.isatra.2020.12.054>
- Wu, X., Zhang, Y., Cheng, C., & Peng, Z. (2021). A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. *Mechanical Systems and Signal Processing*, 149, Article 107327. <https://doi.org/10.1016/j.ymssp.2020.107327>
- Xiao, D., Qin, C., Yu, H., Huang, Y., Liu, C., & Zhang, J. (2021). Unsupervised machine fault diagnosis for noisy domain adaptation using marginal denoising autoencoder based on acoustic signals. *Measurement: Journal of the International Measurement Confederation*, 176(February), Article 109186. <https://doi.org/10.1016/j.measurement.2021.109186>
- Xu, J., Ma, L., Zhang, W., Yang, Q., Li, X., & Liu, S. (2019). An Improved Hybrid CNN-SVM based Method for Bearing Fault Diagnosis Under Noisy Environment. *Proceedings of the 31st Chinese Control and Decision Conference, CCDC 2019, 2018*, 4660–4665. <https://doi.org/10.1109/CCDC.2019.8832683>.
- Yang, D., Karimi, H. R., & Sun, K. (2021). Residual wide-kernel deep convolutional auto-encoder for intelligent rotating machinery fault diagnosis with limited samples. *Neural Networks*, 141, 133–144. <https://doi.org/10.1016/j.neunet.2021.04.003>
- Yang, J., Bao, W., Liu, Y., Li, X., Wang, J., Niu, Y., & Li, J. (2021). Joint pairwise graph embedded sparse deep belief network for fault diagnosis. *Engineering Applications of Artificial Intelligence*, 99(61627901), Article 104149. <https://doi.org/10.1016/j.engappai.2020.104149>
- Yao, L., Fang, Z., Xiao, Y., Hou, J., & Fu, Z. (2021). An intelligent fault diagnosis method for lithium battery systems based on grid search support vector machine. *Energy*, 214, Article 118866. <https://doi.org/10.1016/j.energy.2020.118866>
- Zhang, K., Lin, N., Tian, G., Yang, J., Wang, D., & Jin, Z. (2022). Unsupervised-learning based self-organizing neural network using multi-component seismic data: Application to Xujiahe tight-sand gas reservoir in China. *Journal of Petroleum Science and Engineering*, 209(November 2021), Article 109964. <https://doi.org/10.1016/j.petrol.2021.109964>