OXFORD CDE

Journal of Computational Design and Engineering, 2023, **10**, 1804–1820

DOI: 10.1093/jcde/qwad076

Advance access publication date: 31 July 2023

Research Article

Multi-head de-noising autoencoder-based multi-task model for fault diagnosis of rolling element bearings under various speed conditions

Jongmin Park[1], Jinoh Yoo[1], Taehyung Kim[1], Jong Moon Ha[2,*] and Byeng D. Youn[1,3,4,*]

[1]Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul 08826, Republic of Korea
[2]Intelligent Wave Engineering Team, Korea Research Institute of Standards and Science (KRISS), Daejeon 34113, Republic of Korea
[3]Institute of Advanced Machines and Design, Seoul National University, Seoul 08826, Republic of Korea
[4]OnePredict Inc., Seoul 06160, Republic of Korea
*Correspondence: jmha@kriss.re.kr (J.M.); bdyoun@snu.ac.kr (B.D.Y.)

Abstract

Fault diagnosis of rolling element bearings (REBs), one type of essential mechanical element, has been actively researched; recent research has focused on the use of deep-learning-based approaches. However, conventional deep-learning-based fault-diagnosis approaches are vulnerable to various operating speeds, which greatly affect the vibration characteristics of the system studied. To solve this problem, previous deep-learning-based studies have usually been carried out by increasing the complexity of the model or diversifying the task of the model. Still, limitations remain because the reason of increasing complexity is unclear and the roles of multiple tasks are not well-defined. Therefore, this study proposes a multi-head de-noising autoencoder-based multi-task model for robust diagnosis of REBs under various speed conditions. The proposed model employs a multi-head de-noising autoencoder and multi-task learning strategy to robustly extract features under various speed conditions, while effectively disentangling the speed- and fault-related information. In this research, we evaluate the proposed method using the signals measured from bearing experiments under various speed conditions. The results of the evaluation study show that the proposed method outperformed conventional methods, especially when the training and test datasets have large discrepancies in their operating conditions.

Keywords: fault diagnosis, various speed conditions, de-noising autoencoder, multi-head CNN, multi-task learning

List of symbols

f_θ :	Encoder function of an autoencoder
g_ϕ :	Decoder function of an autoencoder
$x^{(i)}$:	Original training data sample
$\tilde{x}^{(i)}$:	Noise-added training data sample
$Z^{(i)}$:	Features extracted by the encoder function from input data $\tilde{x}^{(i)}$
$\hat{y}^{(i)}$:	Output of Autoencoder
$\tilde{y}^{(i)}$:	Output of De-noising Autoencoder
N :	Number of training data
$\mathcal{L}_{\text{total}}$:	Total loss function
$\mathcal{L}_k$:	kth loss function of the total
λ_k :	Weight of the kth loss function
$\mathcal{L}_{\text{MDAM}}$:	Total loss function of MDAM
$\mathcal{L}_{\text{AE}}$:	Autoencoder loss function
$\mathcal{L}_{\text{Fault}}$:	Loss function for the fault diagnosis
$\mathcal{L}_{\text{Speed}}$:	Loss function for the speed identification
C :	The number of class for the cross-entropy loss
L_{ic} :	The label of ith training sample with the cth class
P_{ic} :	SoftMax output of the model of ith training sample

1. Introduction

Rolling element bearings (REBs) are used as essential mechanical components in various industrial facilities, such as wind turbines, spindles, and industrial robots (Kim *et al.*, 2022). However, REBs are one of the most frequently faulty mechanical components; when a fault occurs, the fault can cause significant economic loss (Cerrada *et al.*, 2018; Chen *et al.*, 2020; Liu *et al.*, 2021). Therefore, it is crucial to detect bearing faults accurately and in a timely manner. For this purpose, the most conventional approach is to extract fault-sensitive features from vibration signals in various domains, such as the time (Jin *et al.*, 2014; Nayana & Geethanjali, 2017; Nikula *et al.*, 2020; Oh *et al.*, 2022), frequency (Hasan *et al.*, 2019; Rauber *et al.*, 2015; Tyagi & Panigrahi, 2017), time–frequency (Alabsi *et al.*, 2021; Guo & Tse, 2013), and cepstrum (Jiang *et al.*, 2019; Kim *et al.*, 2023a) domains. In addition, various studies related to feature selection and diagnosis using various extracted features have been developed (Oh *et al.*, 2022; Raouf *et al.*, 2022). However, feature-based fault-diagnosis (FD) approaches require significant effort to develop an optimal feature-engineering

Received: February 23, 2023. **Revised:** July 17, 2023. **Accepted:** July 17, 2023
© The Author(s) 2023. Published by Oxford University Press on behalf of the Society for Computational Design and Engineering. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

technique. In order to overcome the limitations of the feature-based approach, many studies have recently been conducted to propose solutions that employ a deep-learning model that enables end-to-end FD with automatically defined fault-related features, without the need for a feature engineering process. Among various approaches, one-dimensional (1D) convolutional neural network (CNN)-based models have been widely studied and proposed for REB FD (Chen *et al.*, 2018; Kim & Youn, 2019; Liu *et al.*, 2020; Peng *et al.*, 2020; Wang *et al.*, 2019).

Despite the promising performance of 1D CNN models, they remain vulnerable to the various speed conditions of the system. Even if the speed does not change time-varyingly, various speed conditions can significantly affect the characteristics of vibration for rotating machinery. This variability can make it challenging to extract only fault-related features using a typical simple CNN network. If speed variability becomes extreme, it can be treated as a domain shift problem (Ding *et al.*, 2023; Yu *et al.*, 2023). Several researches for FD or reliability have proposed using meta-learning or transfer learning to solve domain shift problems (Lin *et al.*, 2022; Yao & Han, 2023). Conventional studies for domain shift problems usually require additional training with target domain data to decrease domain discrepancy. However, suppose the causes of domain difference can be separated from and focused on the information of interest only with the data in the source domain. In that case, it will help solve the domain shift problem without additional training with target domain data. Therefore, the effective approach to disentangle the speed-related information, the cause of domain shift, from the fault-related information can be helpful without additional training.

To solve speed variation challenge without additional training, multi-head (MH) and multi-scale CNN-based models have been proposed (Huang *et al.*, 2022; Kim *et al.*, 2023b; Qiao *et al.*, 2019; Shi *et al.*, 2021). MH CNN is generally configured to employ multiple 1D CNN networks in parallel to increase the model's complexity; this enables the model to extract informative features from a wide range of speed conditions. The features extracted from each of the multiple networks can be automatically synthesized in the latent space to achieve optimal FD performance, even under various speed conditions. To this end, Chen *et al.* (2021a) proposed a bearing diagnosis method that includes a multi-scale CNN-based feature extractor; this approach can be considered a variant of the MH CNN approach, with two different kernel lengths. In this approach, it is expected that each head with a different kernel length is capable of extracting features under different ranges of speed conditions. Qiao *et al.* (2019) proposed a novel model that is configured with various modules with multi-scale feature learning, multi-scale feature weighting, and multi-scale feature fusion. Also, Huang *et al.* (2022) proposed multi-scale CNN with three different pooling layers in parallel, and fused features are used for classification under various speed conditions. In research in another field, Ternes *et al.* (2022) proposed a multi-encoder variational autoencoder (AE) to extract different types of information in a single-cell image. Each of these multi-encoders enables effective extraction of independent features from the data, which have been entangled and are not separated by conventional single-encoder-based models. However, most recent studies on MH CNN-based models do not assign specified roles to each head; instead, they expect that informative features can be automatically extracted. Thus, in some cases, features from each head could be highly correlated, which means that the MH structure can be redundant, without any unique role for each head.

Meanwhile, multi-task learning, a learning method that allows a model to perform multiple specific roles simultaneously, can also solve the challenges that arise from various speed conditions (Wang *et al.*, 2022). For REB diagnosis under various speeds, a simple auxiliary task can be designed to classify various load and speed conditions, as well as to determine the fault mode, using shared parameters (Wang *et al.*, 2022; Xie *et al.*, 2022; Zhao *et al.*, 2021). In details, Xie *et al.* (2022) proposed a new model called MTAGN constructed with two different task identifying networks with attention networks branched from task sharing network for FD and severity identification. However, multi-task approaches, including the research mentioned above, perform multiple tasks from common latent features without considering the entangled information that is related to both the fault and operating conditions in the data. To solve this challenge, Huang *et al.* (2022) proposed a data-driven fault feature separation method (DFSM) that employs an MH and multi-task model that enables FD under varying working conditions. In DFSM, the health-related and working-condition-related information are disentangled through the use of two encoders, and based on the assumption that the fault information is irrelevant to the working conditions. For this purpose, multiple loss functions are employed; classification loss is used to classify the health condition, reference loss is employed to ensure similar working-condition-related features are found from normal and faulty conditions; uncorrelated loss is used to impose independence between the classification and reference losses; and, AE loss proposed to reconstruct the original signal using the health-related and working-condition-related features. However, while this approach optimizes the total loss function, working condition features may easily converge into a zero vector due to uncorrelated loss and reference loss characteristics. This is because if the working condition feature converges to zero, the discrepancy between (i) the normal and fault and (ii) an inner product with a fault feature will go to zero. In addition, the definition of the working condition feature, independent of fault information and common in all health states, is so vague that it cannot reflect the characteristics of the various speed conditions.

In summary, the existing approaches, without requiring additional training to overcome domain discrepancies, are proposed with increased model complexity or multiple involved tasks. However, the combination of these two approaches is rare, and even if present, there are limitations in training additional roles. Therefore, this paper proposes a multi-head de-noising autoencoder-based multi-task (MDAM) model for FD under various speed conditions. MDAM employs a de-noising autoencoder (DAE)-based architecture composed of two CNN-based encoders and one decoder. The AE trained by a multi-task learning strategy enables the extraction of disentangled features from the health states and speed conditions. The features extracted by two encoders are used to reconstruct the input signal provided by the decoder to ensure abundant information. To validate the performance of the proposed method, comparative studies with conventional MH CNN-based approaches and ablation studies are presented. These validations are carried out under two cases: internal and external test conditions, depending on whether the test datasets are measured from inside or outside the operating conditions of the training dataset. For all cases, training and testing conditions are designed so that the operating speeds for each condition do not overlap.

The main contributions of the research outlined in this paper are summarized as follows:

(i) The proposed model is configured as an MH DAE that extracts speed- and fault-related features for robust FD under untrained speed conditions.
(ii) Through multi-task learning, FD performance is improved by independently handling different types of information embedded in the vibration signal of a REB.

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

(iii) The proposed MDAM approach shows good performance and robustness against noise, even when the sparse training speed conditions.

The contents of the remainder of this paper are as follows. In Section 2, we present background knowledge related to the proposed method. After describing MDAM in Section 3, experimental validation is outlined in Section 4. Finally, a conclusion is provided in Section 5.

2. Background Knowledge

This section provides the background knowledge needed to understand the main parts of the proposed method. The proposed model is configured based on a DAE and trained using a multi-task-learning strategy. DAEs are explained in Section 2.1; the MH structure and multi-task learning that constitutes the basis of MDAM are described in Section 2.2.

2.1. De-noising autoencoder

AEs are one of the most well-known neural-network-based feature extraction models with a symmetric architecture. AEs consist of an encoder, a feature extraction part, and a decoder for the reconstruction of the input data using the extracted features (Goodfellow *et al.*, 2016). Usually, AEs are used for unsupervised learning for representative feature extraction or dimension reduction of unlabeled data by the bottle-neck-shaped encoder and it can be expressed as

$$Z^{(i)} = f_\theta^{AE}(x^{(i)}) \tag{1}$$

$$\hat{y}^{(i)} = g_\phi^{AE}(Z^{(i)}) \tag{2}$$

where $Z^{(i)}$ is the feature extracted by encoder of AE (f_θ^{AE}) using ith input data $x^{(i)}$. And the decoder of AE (g_ϕ^{AE}) generates $\hat{y}^{(i)}$ which is similar to input data.

However, in the real world, clean data without any noise are hard to get, and input data can be easily corrupted. To enhance the robustness of the model against contamination of the input data by noise, Vincent *et al.* (2008) proposed a method called a DAE. A DAE is trained to extract features from deliberately corrupted input data, while reconstructing the noise-canceled original data to have noise robustness, specifically:

$$\tilde{Z}^{(i)} = f_\theta^{DAE}(\tilde{x}^{(i)}) \tag{3}$$

$$\tilde{y}^{(i)} = g_\phi^{DAE}(\tilde{Z}^{(i)}) \tag{4}$$

where $\tilde{Z}^{(i)}$ is the feature extracted by the encoder function of DAE (f_θ^{DAE}) from ith noise added to input data $\tilde{x}^{(i)}$. $\tilde{y}^{(i)}$ represents the output of the decoder (g_ϕ^{DAE}), which is reconstructed from the feature $Z^{(i)}$. The AE constructed in equations (1) and (2) is trained to extract the key features that can produce an output that is as similar to the input as possible, through the loss function expressed as

$$L_{AE} = \frac{1}{N}\sum_{i=1}^{N} \left\| x^{(i)} - \hat{y}^{(i)} \right\|_2 . \tag{5}$$

Latent features extracted through the loss function are related to the primary information of the data. In particular, for FD, many FD studies have been conducted using such extracted features, as these features contain a lot of information related to faults (Jang *et al.*, 2023; Lu *et al.*, 2017a; Wu *et al.*, 2021).

2.2. MH structure and multi-task learning

In a conventional CNN structure, convolution layers are usually serially configured; the single convolution operator in each layer is trained to extract features from information handed from previous layers in the whole network. Thus, the deep layer creates the higher level feature through a concept extracted from the preceding layers (Lin *et al.*, 2014). Several convolution operators positioned in parallel can strengthen the diversity of features extracted in a single layer. For more complex and diverse feature characteristics, the lengths of kernels in single layers can be varied; this is called multi-scale CNN (Chen *et al.*, 2021b). However, due to the limitations of serially configured CNN structures, features have to be extracted through shared parameters. This phenomenon inevitably hinders the model's ability to extract diverse features and carry out various tasks independently; thus, it is better to proceed with feature extraction through a separate network. Therefore, independent MH networks have been proposed to extract various meaningful features (Li *et al.*, 2020; Ternes *et al.*, 2022).

Usually, a model with learnable parameters is trained to perform only one task and the task that the model performs is expressed as a loss function, such as classification or regression. However, the situation where a model is expected to perform several tasks simultaneously can be plausible. A multi-task learning strategy can be employed when a model is asked to carry out several tasks simultaneously through the use of multiple loss functions. To train these various tasks at the same time, a total loss function is constructed as a linear combination of each loss function corresponding to each task to be trained, as

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_1 + \lambda_2\mathcal{L}_2 + \ldots + \lambda_n\mathcal{L}_n = \sum_{k=1}^{n} \lambda_k\mathcal{L}_k \tag{6}$$

where n is the number of tasks to be trained simultaneously, $\mathcal{L}_k$ is the loss function of the kth task, and λ_k means the weight of each task. The weight of each loss function can be proportional to the importance of each task; however, if there are no special differences between any other sets of λ_k, all λ_k, the values are usually set as one (Li *et al.*, 2020; Wang *et al.*, 2022).

3. Proposed Method

This section describes the proposed MDAM model. Sections 3.1 and 3.2 introduce the motivation and provide a detailed explanation of the proposed method, respectively. Section 3.3 outlines the strategy used to train the proposed MDAM model and provides the overall flow chart for the training and testing.

3.1. Motivation of the proposed method

A fault signal of a REB has three types of information, i.e., fault-related, speed-related, and noise (Antoni & Randall, 2003). If the operating condition (e.g., rotational speed) is fixed, a CNN-based model (i.e., DAE) can extract the fault-related features for robust FD, even under severe noise conditions (Lu *et al.*,). These types of information are entangled, so changing one type can affect the other type of information. Various speeds, which are common in real-world settings, can significantly shift the feature distribution and make it hard to diagnose the fault. To address this issue, and achieve robust FD performance, it is necessary to separate the fault-related information from other types of information in the signal.

A diagram of the feature distribution is illustrated in Fig. 1 for detailed expression. As expressed in Fig. 1a, a conventional classification model trained to perform FD may extract well-classified

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

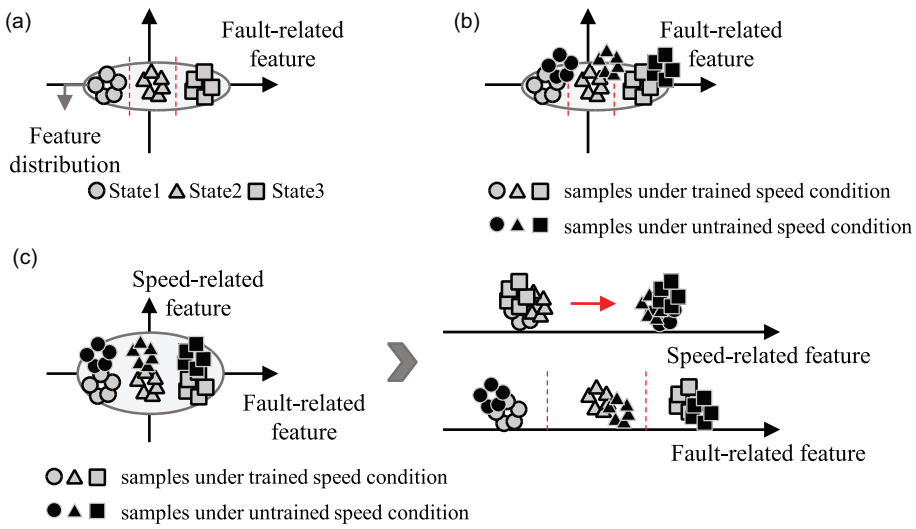

Figure 1: Diagrams of the feature distribution according to each case. (a) Fault-related features with samples under trained speed conditions. (b) Fault-related features with samples under trained and untrained speed conditions. (c) Fault- and speed-related features with samples under trained and untrained speed conditions.

features. Decision boundaries distinguish extracted fault-related features under specific speed conditions; thus, misclassification of samples is rare for single-speed data. However, when the speed information varies, the fault-related features are affected. Because fault features are entangled with speed information, fault-related features under untrained speed conditions can cause misclassification. Therefore, there is a need for a way to control other information and accurately extract fault-related information. As mentioned in the introduction section, some research on training with additional target domain data has been proposed to address this challenge as a domain shift problem. However, separating the cause of domain shift from the entangled information can be helpful without additional training for lessening domain discrepancy. Independently handling speed- and fault-related information alleviates the need for domain adaptation techniques. To this end, this research aims to separately extract the two specified primary types of information (i.e., speed- and fault-related). The goal is to ensure good FD performance by allowing two different features to accurately control the different pieces of information. Thus, the fault-related features are unaffected even if the speed changes, as shown in Fig. 1c.

3.2. MDAM model

As mentioned in Section 3.1, in order to achieve robust FD performance, it is crucial to ensure that the fault-related feature is not affected by the speed-related information. Furthermore, we hypothesize that features extracted from both types of information should contain essential information about the vibration signals, which in turn enables the reconstruction of the input vibration signal through an AE structure. We proposed the MDAM based on an MH multi-task AE network to handle this issue. AE with a single-head (SH) encoder can extract important features to reconstruct data from the extracted features, but uninformative features may be extracted together with the important features. To overcome the aforementioned issue, Ternes *et al.* (2022) proposed an MH AE structure that is not sharing parameters, so the model successfully controls uninformative features independently extracting features. Furthermore, each encoder is forced to be trained to classify two different types of classes independently using multi-task learning, so they can be fitted to different classification goals and finally extract two disentangled features independently. Without a specific task, features that are unclear in meaning but help restore signals can be extracted. Still, each encoder is trained to focus on physically identifiable features to help analyze the extracted features in a well-known way, such as signal processing.

And the details of proposed MDAM model can be represented as shown in Fig. 2. The proposed model is composed of (i) an MH encoder, (ii) a decoder, (iii) speed identification (SI), and (iv) FD. The encoder part was constructed based on an MH model composed in parallel with [Fig. 2 (1–1)] a speed-related encoder and [Fig. 2 (1–2)] a fault-related encoder, unlike conventional approaches. To make these encoders extract robustly independent features for interception of noise, the noise-added data are used as input to perform like DAE as mentioned in Section 2.1. Using the features extracted from the speed-related encoder [Fig. 2 (1–1)], SI is achieved by the classification task [Fig. 2 (3)]. At the same time, FD is achieved through another classification task [Fig. 2 (4)] using the features extracted from the fault-related encoder [Fig. 2 (1–2)]. Thus, it can be seen from Fig. 2 (2) that the decoder combines the fault- and speed-related features to reconstruct the input signal. To improve the feature extraction performance, we added an attention module, as shown in Fig. 3, after the last convolution layer of each encoder branch.

Table 1 describes the detailed structure and hyperparameter values of the proposed MDAM model. The speed- and fault-related encoders are composed of the same structure and hyperparameters. The components of each encoder are convolution blocks consisting of five convolution layers, an exponential linear unit (ELU) activation function, batch normalization, a max pooling layer, an attention module, and a fully connected layer. Actually, ELU activation functions are deliberately utilized for extracting features oscillating around zero in a signal-like way. The filter sizes of the convolution layers are set to 3, and the max pooling layer pool size is set as 2. Attention modules are designed to focus on a certain region of features and ensure higher performance, so convolution layers of attention modules use 1×1 convolutional filter. After three consecutive convolutional operations, feature extrac-

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

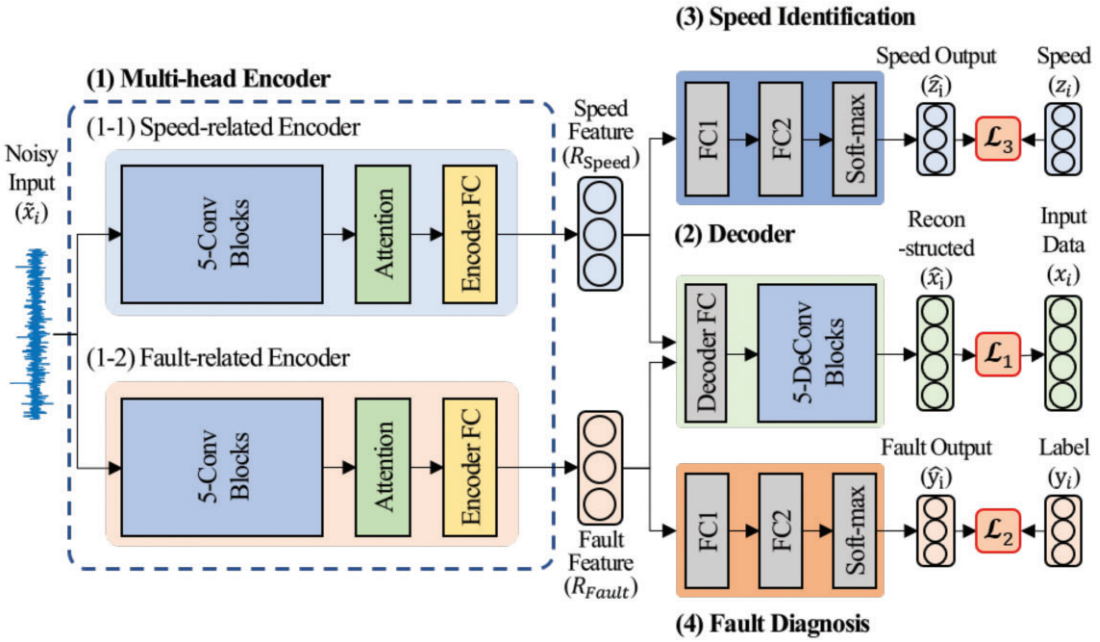

Figure 2: MDAM model configuration.

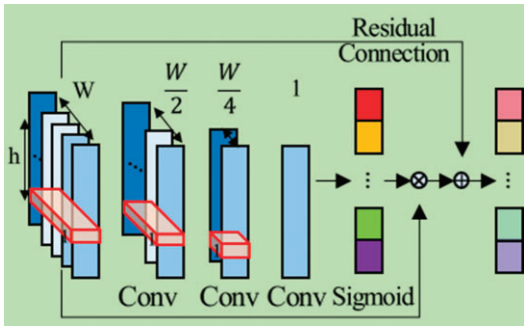

Figure 3: Details of the attention module.

tion is completed using a skip connection with the input signal of the attention modules. Further, to extract features more densely, a fully connected layer is attached after the attention modules. The decoder modules consist of fully connected layers and transposed convolution layers in the reverse order to that of the encoder module. Due to the two types of features in the reconstructing, the number of channels is double that of the encoders. However, the kernel length and upsample size are the same as that of the encoders. Finally, the speed-identification module and FD module are composed of two fully connected layers. The hyperparameters of MDAM, including those mentioned above, were initially set based on conventional methods and later fine-tuned using a heuristic approach rather than an additional optimization algorithm.

3.3. Multi-task learning for MDAM

To train the proposed MDAM model efficiently, we utilize the training strategy called multi-task learning, described in Section 2.2. Because MDAM performs three tasks (i.e., autoencoding, FD, and SI) it learns by optimizing loss functions for each task. Thus, the total loss function of MDAM ($\mathcal{L}_{\text{DAMM}}$) is expressed as

$$\mathcal{L}_{\text{DAMM}} = \mathcal{L}_{\text{AE}} + \mathcal{L}_{\text{Fault}} + \mathcal{L}_{\text{Speed}} \tag{7}$$

where $\mathcal{L}_{\text{AE}}$ is the AE loss for the reconstruction of the signal, $\mathcal{L}_{\text{Fault}}$ is the loss function for the FD, and $\mathcal{L}_{\text{Speed}}$ is the loss function for the SI. The AE loss is defined as the mean square error to represent how the AE reconstructs the signal despite the added noise, which can be expressed as

$$\mathcal{L}_{\text{AE}} = -\frac{1}{n}\sum_{i=1}^{n}\left\|\tilde{x} - \hat{x}\right\|^2. \tag{8}$$

The second and third loss terms of the total loss function are cross-entropy losses for the classification tasks, which can be expressed as

$$-\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{C} L_{ic}\log(P_{ic}). \tag{9}$$

In the equation, n is the number of samples in the training dataset, C is the number of classes; $C = 3$ for FD (i.e., three health classes) and $C = C_s$ for SI, where C_s is the number of speed classes for the training of the model.

Using the total loss function in equation (10), training and testing of MDAM is carried out as detailed in the overall flow chart in Fig. 4. As explained in Section 3.1, in order to evaluate the robustness of the performance of MDAM for datasets with untrained speed conditions, the training and test data are divided based on the speed condition. The model is trained with noise-added signals and two speed- and health-state labels for multi-task learning. The 5 dB level of white Gaussian noise is added to the training data, but the test data are used as raw data. To update the parameters of the MDAM model, gradient-descent-based optimization is conducted using the total loss function. The iteration using these loss functions is continued for the pre-set epoch value. The epoch value is set as 50 and during training ADAM optimizer (Kingma & Ba, 2015) with 0.001 learning rate and 128 size of mini-batch.

After training, the trained MDAM model is used to diagnose the test-speed condition data. The test proceeds without any label information; only vibration data are used to diagnose by MDAM. The noise addition is not also applied; therefore, diagnosis proceeds with only data obtained in a purely experimental environment.

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

Table 1: Detailed structure of the MDAM model.

Module	Block	Layer type	Channel no.	Kernel size	Activation	Stride
Speed & fault encoder	Conv. block1	Convolution	128	3	ELU	1
		Max pool	128	2	-	2
	Conv. block2	Convolution	32	3	ELU	1
		Max pool	32	2	-	2
	Conv. block3	Convolution	16	3	ELU	1
		Max pool	16	2	-	2
	Conv. block4	Convolution	8	3	ELU	1
		Max pool	8	2	-	2
	Conv. block5	Convolution	4	3	ELU	1
		Max pool	4	2	-	2
	Attention block	Convolution	2	1	ReLU	1
		Convolution	1	1	ReLU	1
		Convolution	1	1	Sigmoid	1
	Feature extraction	Fully connected	100	-	ELU	-
Decoder		Fully connected	256		ELU	
		Transposed conv.	128	3	ELU	1
		Transposed conv.	32	3	ELU	1
		Transposed conv.	16	3	ELU	1
		Transposed conv.	8	3	ELU	1
		Transposed conv.	1	3	Sigmoid	1
Fault diagnosis		Fully connected	1	128	ELU	
		Fully connected	1	3	-	
Speed identification		Fully connected	1	128	ELU	
		Fully connected	1	3	-	

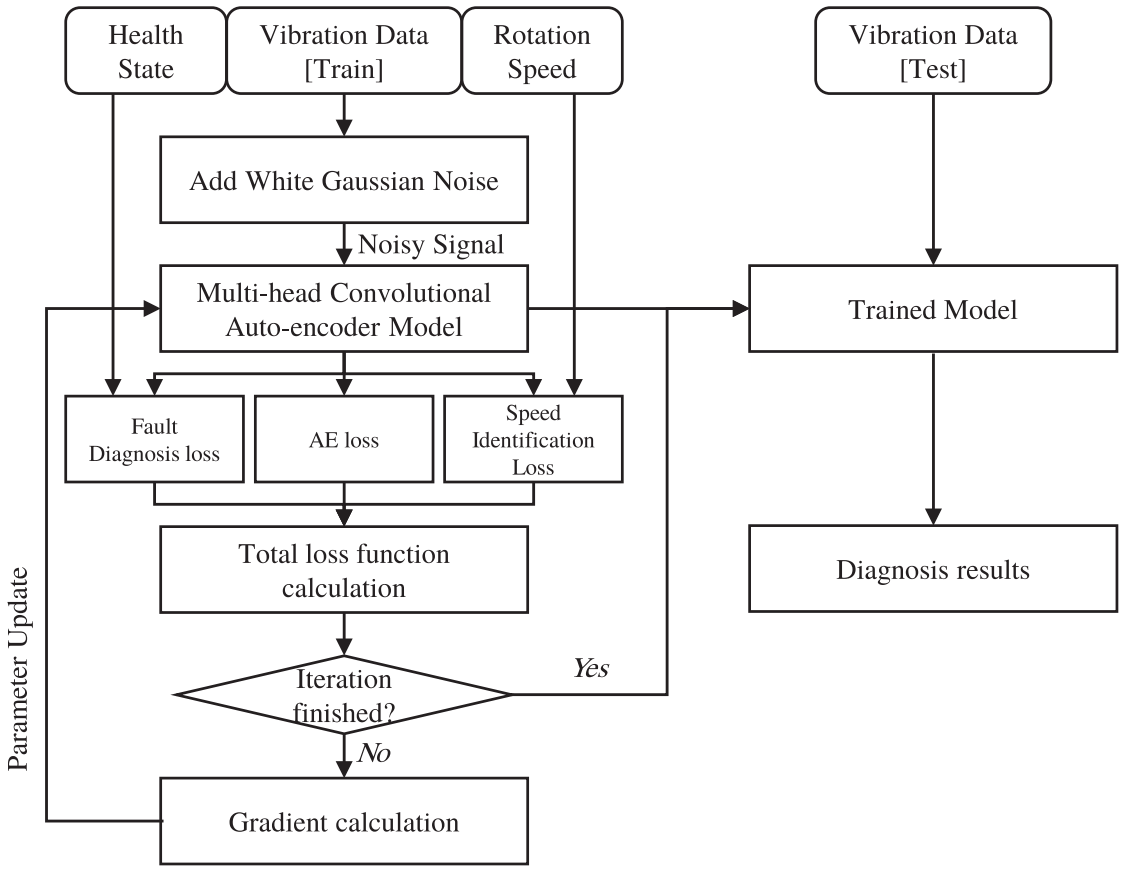

Figure 4: Overall flow chart of the proposed method.

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

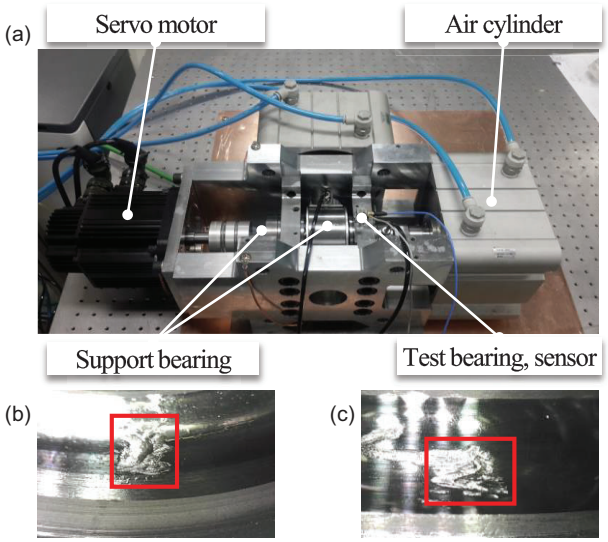

Figure 5: (a) REB testbed rig, (b) inner raceway fault, and (c) outer raceway fault.

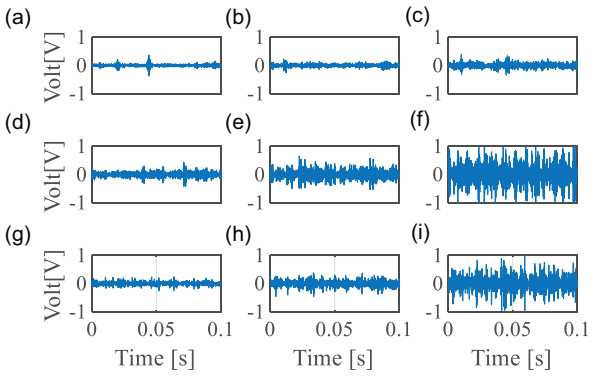

Figure 6: Vibration data. Normal signal under (a) 1000, (b) 1250, and (c) 1500 RPM. Inner race fault signal under (d) 1000, (e) 1250, and (f) 1500 RPM. Outer race fault signal under (g) 1000, (h) 1250, and (i) 1500 RPM.

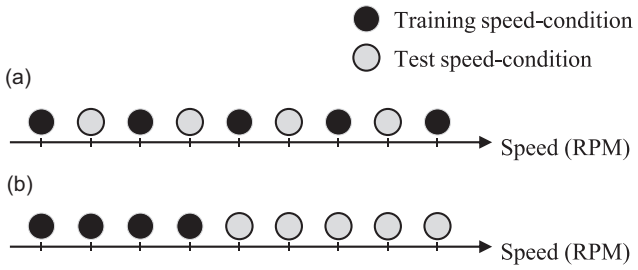

Figure 7: Configuration of the training and test speed conditions. (a) Internal and (b) external case.

4. Experimental Validation

The validation procedure of the proposed MDAM method is presented in Section 4. A detailed description of the REB testbed and the data used to validate the proposed method are provided in Section 4.1. Section 4.2 presents the results for two cases, i.e., testing under internal and external speed conditions. Results are presented for the comparative studies, ablation studies, and noise-robustness tests.

4.1. Description of the REB testbed and data

Vibration signals were acquired from a REB testbed to validate the proposed method, as shown in Fig. 5a. The REB testbed is composed of a servo motor, support bearings, two air cylinders for axial and radial load, and a ball-type test bearing (SKF 7202a). We used a three-axis accelerometer (PCB 356A15) and a National Instrument DAQ module (NI 9234) to acquire vibration data. The sampling frequency for the acquisition of data was set as 10 240 Hz.

Experiments on the REBs were carried out for three types of health states, i.e., normal, inner raceway fault, and outer raceway fault. The two kinds of faults were artificially seeded with a spall-like shape, which is the most common fault type, as shown in Fig. 5b and c. Axial and radial loads were set as 180 and 170 kgf. The signal is acquired for multiple speed conditions to validate the proposed method under various speed conditions. Conventional cross-domain FD research that considers rotational speed as the domain discrepancy factor usually employs two or three speed conditions. For more comprehensive testing of the results, 11 speed conditions ranging from 1000 to 2000 RPM with an interval of 1000 RPM are selected as the experimental speed condition. The testbed was run for 70 sec for each experiment; thus, the number of samples for each dataset is 716 800 (10 240∗70).

Figure 6 shows examples of the data for each condition. The scale of the plots in Fig. 6 is fixed as [−1, 1] for easy data comparison. It can be seen that the rotational speed significantly affects the vibration characteristics, including magnitudes and patterns. Moreover, the rotational speed also considerably affects the characteristics of the fault-related features, so the differences are not well represented in Fig. 6. Thus, if the FD model is trained without considering the speed variation, it will lead to significant misclassification of the health state for the untrained speed condition.

4.2. Validation results

In order to evaluate the speed-invariant performance of MDAM, the validation cases are divided in a detailed way considering following two primary factors: (i) the relation between speeds of training and test data and (ii) the sparsity of training data among whole data speed interval. Firstly, the whole validation case was divided into two cases, specifically, where the test speed condition is placed (i) internally or (ii) externally of the training-speed condition expressed as Fig. 7. Deep-learning models can usually perform well in interpolation case, but they may be vulnerable when they extrapolate (Busemeyer *et al.*, 1997; Sharma & McNeill, 2009). We tested the proposed model not only for easier internal cases but also for more difficult external cases. In addition, the internal and external cases are classified into sub-cases considering the sparsity of the speed condition used for training, as shown in Table 2. The speed interval in the training data controls the sparsity of the test for unseen internal case; from the I-1 sub-case with 200 RPM to I-5 with 1000 RPM. In the external case, the speed interval of the training data was fixed at 100 RPM, while the sparsity was adjusted by the total number of speed conditions used for learning. E-1, which uses the most-abundant speed condition, used seven speeds, from 1000 to 1600 RPM. As the number gradually decreased, E-6 used only two speeds, 1000 and 1100 RPM.

Section 4.2.1 presents the results of the proposed method, along with the results found by the comparative method for both the internal and external cases. The ablation studies for both cases are presented in Section 4.2.2, and the noise-robust tests are given in Section 4.2.3.

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

Table 2: Details of validation cases.

Case	Sub-case	Training data speed (RPM)	Test data speed (RPM)
Test for unseen internal data	I-1	1000, 1200, 1400, 1600, 1800, 2000	1100, 1300, 1500, 1700, 1900
	I-2	1000, 1300, 1600, 1900	1100, 1200, 1400, 1500, 1700, 1800
	I-3	1000, 1400, 1800	1100, 1200, 1300, 1500, 1600, 1700
	I-4	1000, 1500, 2000	1100, 1200, 1300, 1400, 1600, 1700, 1800, 1900
	I-5	1000, 2000	1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900
Test for unseen external data	E-1	1000, 1100, 1200, 1300, 1400, 1500, 1600	1700, 1800, 1900, 2000
	E-2	1000, 1100, 1200, 1300, 1400, 1500	1600, 1700, 1800, 1900, 2000
	E-3	1000, 1100, 1200, 1300, 1400	1500, 1600, 1700, 1800, 1900, 2000
	E-4	1000, 1100, 1200, 1300	1400, 1500, 1600, 1700, 1800, 1900, 2000
	E-5	1000, 1100, 1200	1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000
	E-6	1000, 1100	1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000

4.2.1. *Comparative study*

In the comparative study, we show the results of the proposed approach, compared with previous methods, to demonstrate the performance of the proposed MDAM model. Among previous methods, we selected four representative methods that had been developed to able to cope with various speed conditions by employing an MH or multi-task learning strategy. The first model, called a multi-scale convolutional neural network with channel attention (CA-MCNN, Huang *et al.*, 2022) is compared with the proposed method. The model is constructed with multi-scale feature learning part and multi-scale feature fusion with a weighted fusion method by a convolution operation. The second one is MTAGN, constructed with task-specific attention layers branched from task-a shared network trained by multi-task learning with a special training strategy, dynamic weight average. MTAGN was previously proposed for fault type identification and fault severity identification; however we revised this model slightly for fault type classification and SI in this research. The third comparative model is DFSM (Li *et al.*, 2020), an MH AE-shaped model that extracts working-condition-related and fault-related features under various speed conditions. Finally the last comparative model is AWMSCNN (Qiao *et al.*, 2019), which has an MH CNN classifier using an MH structure that can enable FD under various speed conditions. Because our model uses noise-added data during training, the comparative studies are carried out using the aforementioned comparative models with and without the white Gaussian noise of 5 dB. The noise is added during the training process before being put in each model.

Untrained internal speed condition

The results of comparative methods for an untrained internal speed condition are presented in Fig. 8 and Table 3. In Fig. 8, the legend marked as noise in parentheses shows the results from the noise-added data; other cases show the results for the original data. As seen from the bar graph of Fig. 8, the proposed MDAM method shows higher accuracy for all sub-cases. Both noise-added data and original data cases show good performance, with a minimum accuracy of 99.74%. From I-1 to I-3, the sparsity may not be seen as the primary factor of accuracy, only the AWMSCNN model represents significant performance degradation, and the other four models show high accuracy of over 99%. But the MTAGN and DFSM models show comparatively significant performance degradation from the I-4 sub-case, where the sparsity becomes severe. The CA-MCNN model shows relatively good performance when compared with other comparative models, but the performance of CA-MCNN declines in the I-5 sub-case more than MDAM.

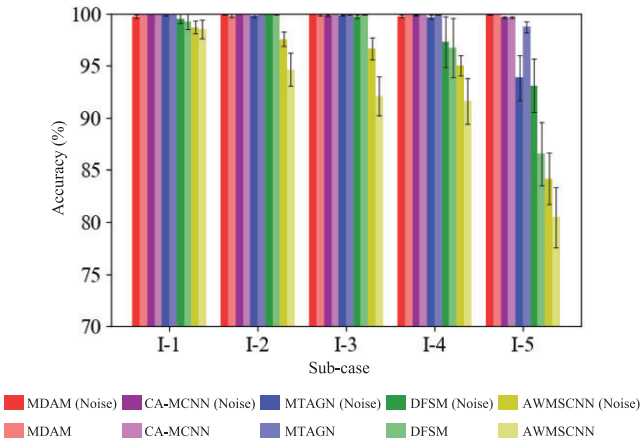

Figure 8: Comparative study results for the untrained internal speed case (bar plot).

In order to examine the comparative study results more deeply, features extracted from the AE and latent features of the hidden layer of the CNN model were plotted and analyzed. The t-distributed stochastic neighbor embedding (t-SNE) method (van der Maaten & Hinton, 2008) was used to represent the latent features of the proposed and comparative models. Figure 9 is a plot of the latent features represented using t-SNE, for the I-1 sub-case in Fig. 9a, and I-5 in Fig. 9b. The features of the proposed MDAM and four comparative study models are presented in each plot. For MDAM, MTAGN, and DFSM, two different kinds of features are presented (i.e., fault feature and speed feature for MDAM and MTAGN, and fault feature and working-condition feature for DFSM). CA-MCNN and AWMSCNN results are presented only for fault features because of the nature of the method. The figure shows following three types of information: health state as the shapes of the markers, speed as the edge color, and training and test dataset distinction as the marker fill style.

The two leftmost figures in Fig. 9a are the representation of features from the proposed MDAM method obtained through t-SNE. It can be seen that the fault feature has a neglectable distribution discrepancy between the training and test datasets across the entire range of speed conditions. All health classes are classifiable; misclassified samples are rare. On the other hand, the speed features appear sequentially according to the speed, while each of them is gathered regardless of the health state. This means that the speed and fault features are effectively disentangled from the feature space by the proposed MH and multi-task model. In addition, it can be seen that the test dataset is distributed between

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

Table 3: Comparative study results for the untrained internal speed case (accuracy table).

Accuracy (%)	Sub-case				
	I-1	I-2	I-3	I-4	I-5
MDAM (noise)	99.74 ± 0.19	**99.97 ± 0.01**	**99.97 ± 0.01**	99.76 ± 0.12	99.91 ± 0.01
MDAM	**100.00 ± 0.00**	99.78 ± 0.15	99.88 ± 0.08	**99.98 ± 0.00**	**99.95 ± 0.01**
CA-MCNN (noise)	99.99 ± 0.00	99.96 ± 0.04	99.96 ± 0.02	99.94 ± 0.02	99.62 ± 0.25
CA-MCNN	99.99 ± 0.01	99.93 ± 0.01	99.84 ± 0.24	99.87 ± 0.06	99.65 ± 0.14
MTAGN (noise)	99.86 ± 0.28	99.82 ± 0.48	99.81 ± 0.13	99.66 ± 0.70	93.85 ± 6.88
MTAGN	99.98 ± 0.02	99.97 ± 0.02	99.90 ± 0.09	99.88 ± 0.07	98.71 ± 1.61
DFSM (noise)	99.52 ± 0.45	99.95 ± 0.01	99.72 ± 0.15	97.31 ± 2.43	93.10 ± 2.53
DFSM	99.23 ± 0.72	99.89 ± 0.04	99.90 ± 0.03	96.74 ± 2.82	86.54 ± 3.07
AWMSCNN (noise)	98.68 ± 0.61	97.54 ± 0.69	96.65 ± 1.03	95.00 ± 0.95	84.17 ± 2.47
AWMSCNN	98.51 ± 0.92	94.63 ± 1.59	92.08 ± 1.88	91.61 ± 2.20	80.45 ± 2.86

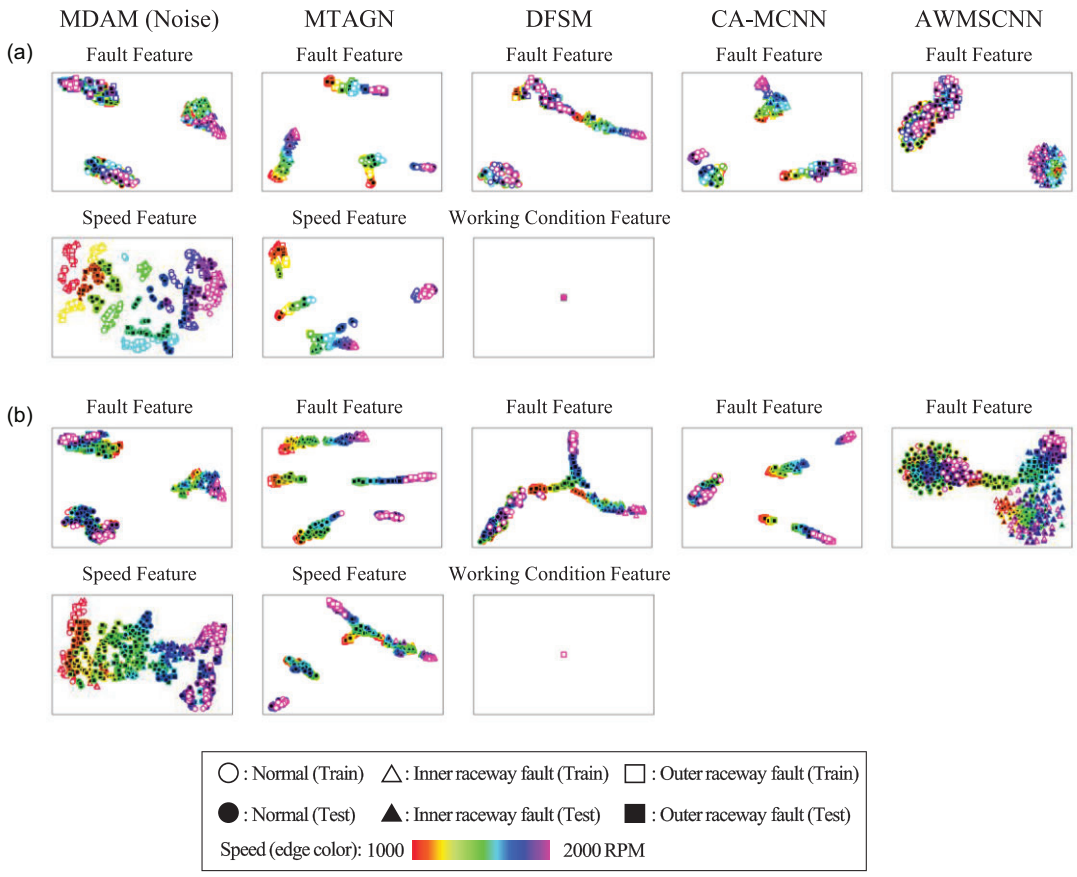

Figure 9: Feature representation for the comparative study to represent speed variety and the difference between the training and test speed conditions. (a) I-1 sub-case and (b) I-5 sub-case.

the distributions of training datasets without losing the physical meaning of the speed condition. For example, features from the test dataset from the 1100 and 1300 RPM speed conditions (i.e., filled markers with the orange and green edges) are located between the features from the training dataset from 1000 and 1400 RPM (i.e., empty markers with the red and green edges). This was not learned in the form of a continuous label, such as regression; however, it can be considered that the speed, which is the physical quantity that the label represents, has been properly trained. In the cases of MTAGN and CA-MCNN, fault features are well-classified; however, the normal samples are separated and not well-clustered. It can be seen that the fault features of DFSM are not well classifiable for each health state. Samples of normal and outer raceway faults are well-clustered, and the variations along the speed change are not severe. However, in the inner raceway fault case, the changes that occur due to speed are so evident that the diagnostic accuracy was harmed. In the case of AWMSCNN, it was found that neither the training nor the test dataset was able to accurately control the speed. Meanwhile, the speed features of MTAGN are separated by health states, and they are not aligned according to speed well. The working condition features show that most are concentrated in one place and converge to near zero. The easiest way to optimize the loss function composed of MMD loss and inner product loss with fault feature is by converging to zero

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

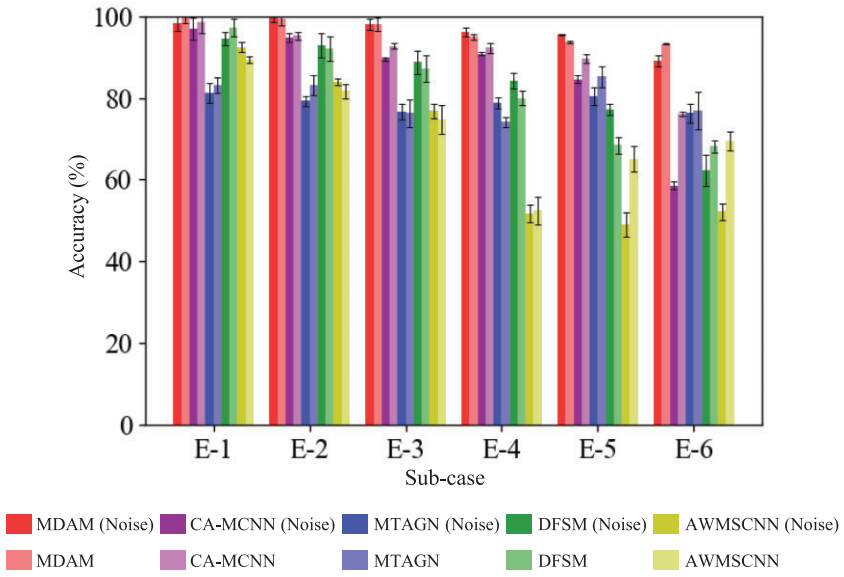

Figure 10: Comparative study results for the untrained external speed case (bar plot).

vector. This is because when the working condition feature goes to zero vector, the inner product loss is zero and the distribution difference along the dataset is also zero. Further, it was difficult to create a classifiable distribution.

Figure 9b shows that the fault feature of MDAM is well-classable, similar to Fig. 9a. In addition, as shown in Fig. 9b, the speed feature of the proposed method represents the physical meaning of the speed condition. The test datasets from 1100 to 1900 RPM speed conditions are sequentially located by the speed order between the training datasets, from 1000 to 2000 RPM speed conditions. The fault features of MTAGN and CA-MCNN and the speed information are not well disentangled, so fault features are aligned along to speed. And the fault features are separated by health state but they are not well sufficiently classified. For DFSM, the training data for all three classes are well classified; however, the deviation that arises due to the speed change is too large to allow the fault features to be well controlled. So, the test data of inner raceway and outer raceway are not well clustered but close other classes, which seemed to have an effect on lowering the classification accuracy. The working condition of DFSM showed the same result with a near-zero vector. The fault feature of AWMSCNN has a small number of speed conditions used for learning, making it difficult to clearly distinguish both the fault feature training and the test dataset.

Untrained external speed condition

The comparative study for this case is shown in Fig. 10 and Table 4, using the same aforementioned process. Even in the case of tests that use external speed condition data, the proposed method shows the highest diagnostic accuracy in all sub-cases. The average accuracy through whole sub-cases relatively lower than internal speed condition, so it is experimentally proved that untrained external speed condition sub-cases are more difficult than internal sub-cases. Furthermore, the sparsity of the training speed condition is more fatal than the internal case, resulting in a greater reduction in accuracy. In the meantime, MDAM still outperforms the comparative methods in terms of accuracy and usually show more than 90%. But CA-MCNN showing the best performance among comparative studies presents relatively low accuracy, specially, 60∼70% in E-6 sub-case. Accuracy of MTAGN which is second accurate model in internal sub-cases is lower than 90% for all sub-cases. Even worse, from E-4 to E-6 sub-cases, fault-related features are not properly controlled to the extent that AWMSCNN shows an accuracy of only about 50%, and the above results can be seen in Fig. 10 and Table 4.

As before, Fig. 11 is expressed by dividing the training/test dataset into different colors according to the speed. Figure 11a and b, respectively, are the results of the E-1 sub-case with the largest training dataset (i.e., 1000 to 1600 RPM speed conditions), and the E-6 sub-case with the smallest training dataset (i.e., 1000 and 1200 RPM speed conditions).

In Fig. 11a, it can be seen that the fault feature of MDAM has a superior consistency between the training and test datasets. The fault feature seems to have higher variability according to speed than observed for the internal case; however, the distributions are well-classified and do not affect the accuracy. On the other hand, the fault feature of MTAGN, DFSM, and CA-MCNN seems to be well classified in the training dataset. Still, it can be seen that the test dataset is plotted in a direction that deviates significantly from each group. The variation in the fault feature of DFSM and MTAGN that appears according to the speed is larger than that for MDAM; thus, the high-speed data are located close to the other classes. The latent features of AWMSCNN are not well clustered in the test dataset; therefore, the overall classification accuracy is low. The speed feature of MDAM was shown to be continuously distributed along all speed conditions for the training dataset. However, because the speed condition for test data is placed externally, the speed features are not well positioned according to speed change. This may be considered as the limitation for extrapolation of the deep-learning model; however, the model can distinguish the difference from train speed condition and also regard test speed as higher than train speed. From these results, it can be concluded that the speed variation from the high-speed condition does not significantly affect the features. The speed features of MTAGN are distributed along the speed value at a glance, but features in the training speed conditions are already classified according to health states. Thus, it can be found from Fig. 10 and Table 4 that all models have relatively high FD performance.

Figure 11b uses relatively fewer speed conditions for training; thus, the difference in the distribution of MDAM's fault features

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

Table 4: Comparative study results for the untrained external speed case (accuracy table).

Accuracy (%)	Sub-case					
	E-1	E-2	E-3	E-4	E-5	E-6
MDAM (noise)	98.36 ± 1.37	**99.53 ± 0.18**	**98.10 ± 1.11**	**96.07 ± 1.35**	**95.44 ± 0.95**	89.12 ± 2.01
MDAM	**99.74 ± 0.17**	99.29 ± 0.39	98.02 ± 0.61	94.93 ± 1.52	93.68 ± 1.52	**93.24 ± 1.56**
CA-MCNN (noise)	98.00 ± 3.57	96.02 ± 0.70	90.16 ± 0.47	90.78 ± 0.47	86.73 ± 0.65	62.34 ± 0.58
CA-MCNN	99.00 ± 2.79	96.59 ± 0.78	94.54 ± 0.73	93.29 ± 0.71	90.75 ± 0.71	72.85 ± 0.19
MTAGN (noise)	81.21 ± 7.34	79.24 ± 7.04	76.68 ± 4.21	78.79 ± 6.05	80.41 ± 3.66	76.30 ± 7.35
MTAGN	83.08 ± 14.72	83.19 ± 8.52	76.26 ± 3.84	74.15 ± 10.69	85.15 ± 7.78	76.90 ± 5.90
DFSM (noise)	94.44 ± 3.80	92.75 ± 1.36	88.71 ± 1.83	84.17 ± 2.86	77.27 ± 2.96	62.36 ± 1.71
DFSM	97.16 ± 1.48	92.08 ± 2.07	87.17 ± 1.84	79.96 ± 3.27	68.38 ± 2.95	68.12 ± 2.21
AWMSCNN (noise)	92.38 ± 1.97	84.02 ± 3.00	76.78 ± 2.22	51.68 ± 1.74	49.04 ± 0.81	52.17 ± 2.16
AWMSCNN	89.43 ± 2.34	81.74 ± 3.02	74.73 ± 3.46	52.42 ± 3.41	65.10 ± 1.76	69.52 ± 0.80

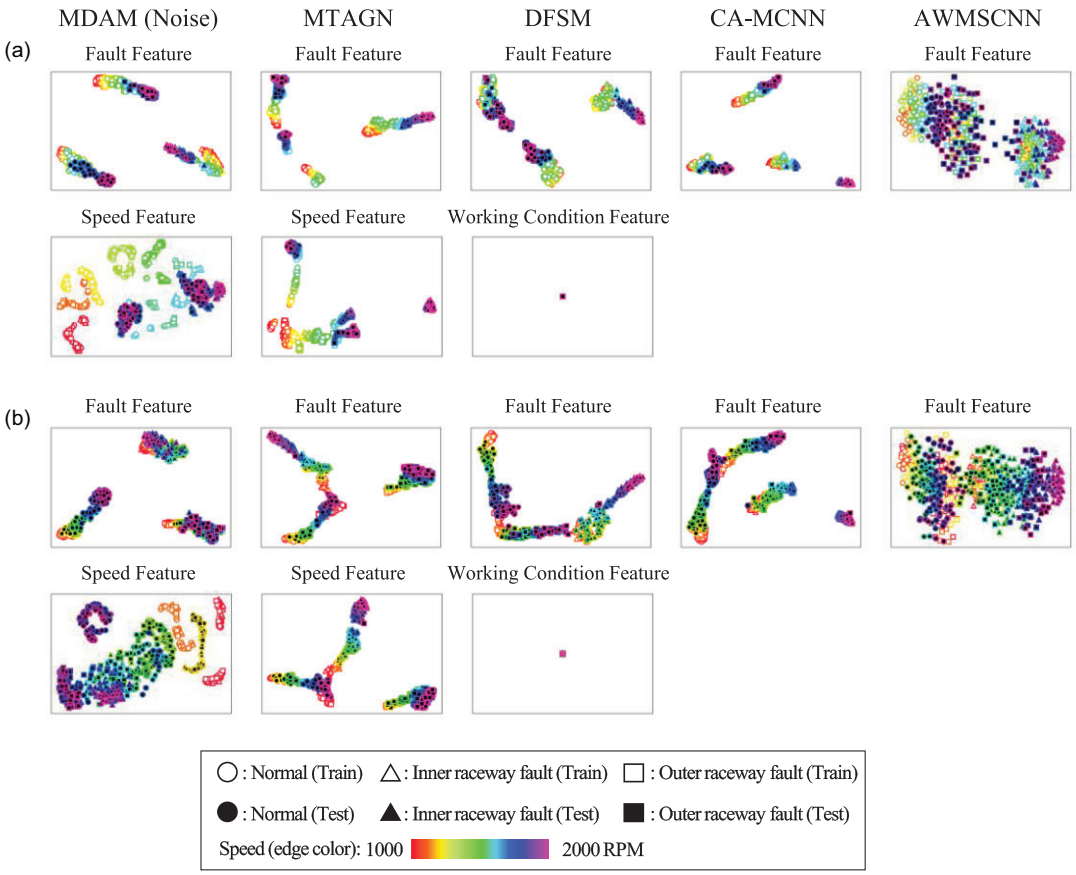

Figure 11: Feature representation for the comparative study to represent speed variety and the difference between the training and test speed conditions. (a) E-1 sub-case and (b) E-6 sub-case.

between the training and test datasets is larger than that observed in Fig. 11a. However, the difference between classes is still clear; therefore, we note that it did not significantly affect the accuracy. In contrast, in the case of the four comparative methods, the speed information could not be accurately controlled; thus, the features were extended to the other class. All comparative models appear to have a test dataset distribution outside the training dataset distribution, and its accuracy decreases because it is in a different class direction. The fault feature for test data of AWMSCNN has a large variation according to speed; therefore, it can be seen that the accuracy is much lower because each class is distributed closer to other health states than to the correct class. AWMSCNN appears to have an effect on accuracy reduction by stretching to other health states. The speed features of MDAM are plotted similarly in Fig. 11a; the continuity is well shown in low-speed data until 1700 RPM but not in high-speed data. The impairment to this continuity means that, here, the fault and speed information are less isolated than in the internal case. When the speed conditions of the test data are located in the training data, speed information is easily inferred by the trained information. However, when the speed of the test data is located externally, it is relatively tricky to handle speed information, and even in the fault feature, variability according to speed was observed. On the contrary, the speed feature of MTAGN is not well controlled, so

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

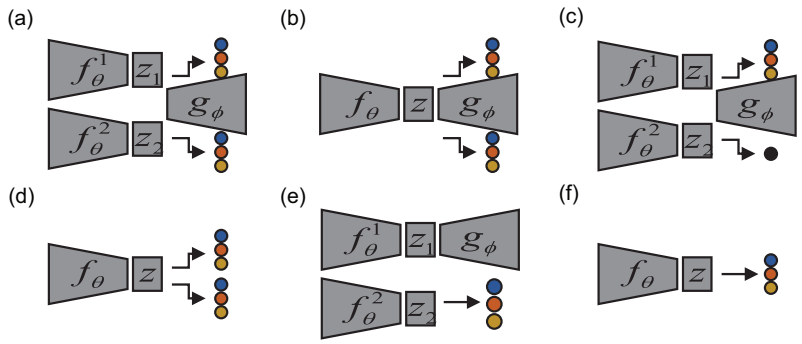

Figure 12: Configurations of the models used in the ablation study. (a) MDAM, (b) SHSIAE, (c) MHSRAE, (d) SHSI, (e) MHAE, and (f) SH.

the trend of the speed feature looks like the fault feature is distributed. Both features are not well separated, and they show similar ways.

4.2.2 *Ablation study*

In the results described in the previous section, the proposed MDAM model showed better results than the comparative models. To further demonstrate its effectiveness, we verified MDAM through an ablation study, which is described in this section. The ablation study determined how much each component of MDAM affects its performance and the overall effectiveness of the proposed method. MDAM is composed mainly of four parts, i.e., MH, FD, SI, and an AE. In order to proceed with the ablation study, comparison of the FD performance is carried out by making models that remove or replace parts to determine the impact of each part. However, since FD is the goal of model development and an indicator of performance comparison, an ablation study was conducted without removing the FD part from the above-described elements. The models used in the ablation study are expressed in Fig. 12 and are described below (FD is excluded because it is common to all models):

(i) Panel a: MH + SI + AE (i.e., proposed MDAM): MH AE-based multi-task FD model with SI.
(ii) Panel b: SH + SI + AE (SHSIAE): SH AE-based multi-task FD model with SI, deleting the structure of extracting different features through an MH structure and extracting features with only one encoder.
(iii) Panel c: MH + SR + AE (MHSRAE): MH AE-based multi-task FD model with speed regression (SR), using the regression task Mean Absolute Error (MAE loss) instead of the classification task in the SI part.
(iv) Panel d: SH + SI (SHSI): SH multi-task FD model with SI to ablate the AE structure. When the AE part is removed, the concatenating parts disappear. Therefore, extracting different features through multi-encoders is meaningless, so SH encoders were used together.
(v) Panel e: MH + AE (MHAE): MH AE-based FD model without the SI part.
(vi) Panel f: SH: SH FD model, similar to a vanilla 1D CNN model without any other parts.

Untrained internal speed condition

Figure 13 and Table 5 show the results of the ablation study conducted for the unseen internal data case test. Similar to the comparative study, all models were trained through data gathered under the same conditions and verified through training with internal speed data. The MDAM model represented the best performance in all sub-cases, showing an average of 99.86% of diagnosis accuracy. In the case of MHSRAE, a model using regression instead of classification of speed, and the SHSI model in which the AE was removed, most situations showed lower accuracy. Unlike the classification task, which requires guessing the highest probable class to be matched from among several classes, the regression task is much more difficult to learn because the regression task requires fitting correct real value. Therefore, the speed encoder of MHSRAE could not handle speed information well, and the diagnostic accuracy was severely diminished. Due to the characteristics of the AE to extract the primary features of the data, it showed accuracy of a little less than 98%, even if the SI part did not specifically guide the encoder of the MHAE, other than the fault encoder. In the case of SH, only one fault-diagnostic task was learned without any other loss function, and the test data also lie in the distribution of the training dataset, indicating that the performance was better than that of other models. However, the performance of MDAM – which offers additional parts capable of controlling speed – was superior.

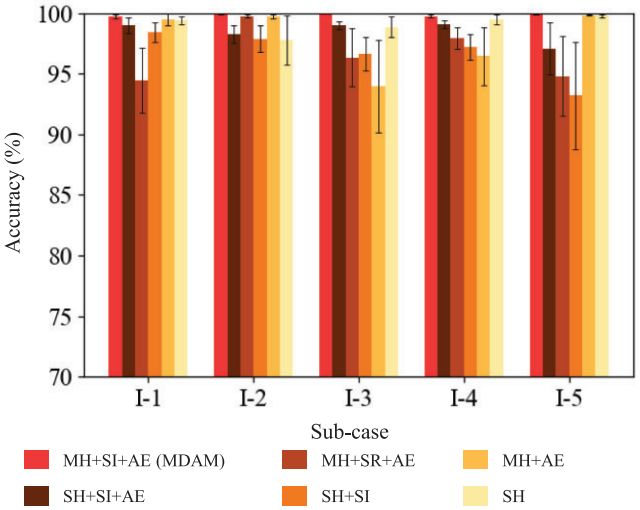

Figure 13: Ablation study results for the untrained internal speed case (bar plot).

Untrained external speed condition

Similar to the internal case, the validity of the proposed method was verified through an ablation study for the external case. The results of the ablation study of the test for unseen external data cases are represented in Fig. 14 and Table 6. MDAM in all sub-cases showed the highest accuracy, with an average accuracy of 96.1%; its difference in accuracy from the other models examined

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

Table 5: Ablation study results for the untrained internal speed case (accuracy table).

Accuracy (%)	Sub-case					
	I-1	I-2	I-3	I-4	I-5	Avg.
MDAM	**99.74 ± 0.19**	**99.93 ± 0.05**	**99.97 ± 0.01**	**99.76 ± 0.12**	**99.91 ± 0.01**	**99.86**
SH + SI + AE	98.98 ± 0.65	98.25 ± 0.72	98.98 ± 0.35	99.05 ± 0.33	97.07 ± 2.12	98.47
MH + SR + AE	94.46 ± 2.66	99.74 ± 0.12	96.34 ± 2.37	97.92 ± 0.88	94.80 ± 3.18	96.65
SH + SI	98.42 ± 0.82	97.89 ± 1.08	96.63 ± 1.39	97.19 ± 1.04	93.21 ± 4.43	96.67
MH + AE	99.44 ± 0.48	99.74 ± 0.15	93.97 ± 3.82	96.46 ± 2.38	99.84 ± 0.04	97.89
SH	99.36 ± 0.33	97.79 ± 2.03	98.85 ± 0.86	99.45 ± 0.41	99.77 ± 0.10	99.04

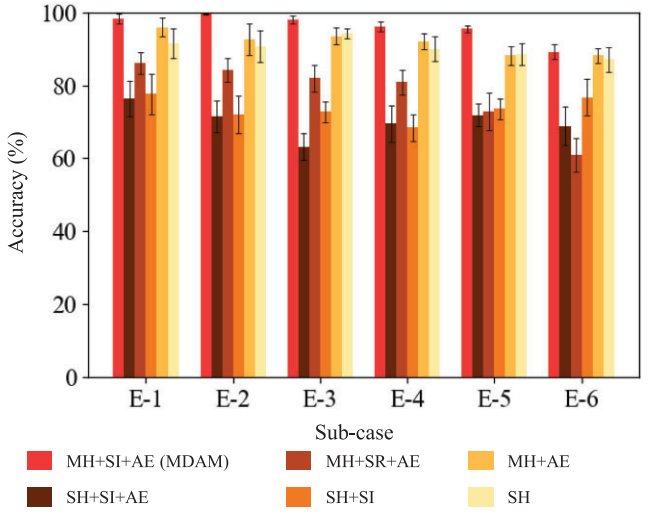

Figure 14: Ablation study results for the untrained external speed case (bar plot).

was larger than that observed for the results of the internal case. The three models – SHSIAE, MHSRAE, and SHSI – show 70% accuracy, on average, indicating that the change in speed greatly affected the diagnosis accuracy. SHSI and SH showed better accuracy than other ablation study models; however, the accuracy was much lower than the internal case, showing an accuracy of just over 90%. In addition, as the sparsity of the training data speed condition increases, the amount of the decrease in accuracy increases. In particular, the accuracy of the three models with low accuracy decreases significantly. However, in the case of MDAM, the E-5 sub-case – where the accuracy of all other models was less than 90% – also exhibited robust FD performance with a high accuracy of 95.44%. However, in the E-6 sub-case, the MDAM model showed a little less than 90% accuracy; however, in this sub-case, it still showed higher accuracy and less variance than the other models.

4.2.3 *Validation for robustness against noise*

In order to validate the noise robustness of the proposed method, comparative validation was conducted with all methods used for comparisons in the previous sections. To evaluate accuracy, white Gaussian noise at various Signal to Noise Ratio (SNR) levels from −5 to 10 was added to the test data for each sub-case. In this test, four comparative models (CA-MCNN, MTAGN, DFSM, and AWM-SCNN) were trained with noise-added data and the models used in the ablation study were compared with MDAM in terms of diagnostic accuracy. In order to validate the performance of the denoising task in MDAM, MDAM was compared both when trained with original data and when trained through the noise-added signal. The accuracy is represented with a bar plot in Figs 15 and 16 with the same colors as used in Fig. 8.

Untrained internal speed condition

In the case of SNRs greater than or equal to 0, good diagnosis accuracy was found in most models and sub-cases; the results are similar to those seen in Figs 8 and Fig. 13. In particular, MDAM showed the best performance, with nearly 100% diagnosis accuracy, both in the case of training with the original data and in the case of training with the noise-added data, showing the robustness of diagnosis for weak noise. On the other hand, when the noise is strong under the SNR condition of −5 dB, the accuracy is lower than in other cases; most accuracy results are 80∼90%. For the I-1 and I-2 sub-cases, other comparative models show higher accuracy than MDAM, such as MTAGN and AWMSCNN. But the MDAM for the sparsest three sub-cases shows the highest accuracy, and the accuracy drop due to the sparsity of training data is less than any other methods. However, the proposed method trained through noise-added data showed better performance than other models, especially when trained with original data.

Untrained external speed condition

In the external case, noise-added data was also used to test various SNR levels. MDAM had the highest accuracy in these tests for most sub-cases and SNR levels. However, the accuracy difference for greater than or equal to 0 SNR between MDAM and other models was greater than that observed in the internal case. Only MDAM shows over 90% accuracy, and other models declined by under 50%. And the training speed sparsity may affect other comparative methods, but the MDAM has represented a smaller accuracy drop due to sparsity. However, for less than 0 SNR level, some models show higher accuracy than MDAM, but the average accuracy is higher than any other methods.

4.2.4 *Feature analysis in frequency domain*

The analysis of hidden layer activations provides an indirect understanding of the functioning of the trained model. Instead of exploring the relationships among samples, such as t-SNE, this section focuses on analyzing hidden layer activations at the individual sample level. To delve deeper into the hidden layer activations, we conduct feature analysis from a signal processing perspective. Given that faulty REB generates cyclic impulses with a period determined by the fault type and geometric information of the REB, the fault characteristics frequencies can indicate the fault type (Randall & Antoni, 2011). By examining the hidden layer activations using fault frequency identification, we can gain insights into how the trained model functions. The MDAM model is designed with an MH to disentangle two types of information, and as such, the different frequencies should be captured by each

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

Table 6: Ablation study results for the untrained external speed case (accuracy table).

Accuracy (%)	Sub-case						
	E-1	E-2	E-3	E-4	E-5	E-6	Avg.
MDAM	**98.36 ± 1.37**	**99.53 ± 0.18**	**98.10 ± 1.11**	**96.07 ± 1.35**	**95.44 ± 0.95**	**89.12 ± 2.01**	**96.10**
SH + SI + AE	76.37 ± 4.92	71.46 ± 4.32	63.21 ± 3.63	69.47 ± 5.02	71.88 ± 3.10	68.87 ± 5.29	70.21
MH + SR + AE	86.05 ± 2.90	84.18 ± 3.19	81.90 ± 3.71	80.83 ± 3.32	72.88 ± 5.16	60.92 ± 4.70	77.79
SH + SI	77.58 ± 5.49	72.02 ± 5.02	72.77 ± 2.83	68.39 ± 3.67	73.58 ± 2.88	76.71 ± 4.92	73.51
MH + AE	95.92 ± 2.54	92.48 ± 4.33	93.49 ± 2.22	91.95 ± 2.15	88.22 ± 2.54	88.16 ± 2.07	91.70
SH	91.48 ± 3.97	90.75 ± 4.32	94.15 ± 1.41	90.00 ± 3.41	88.54 ± 3.04	87.10 ± 3.41	90.34

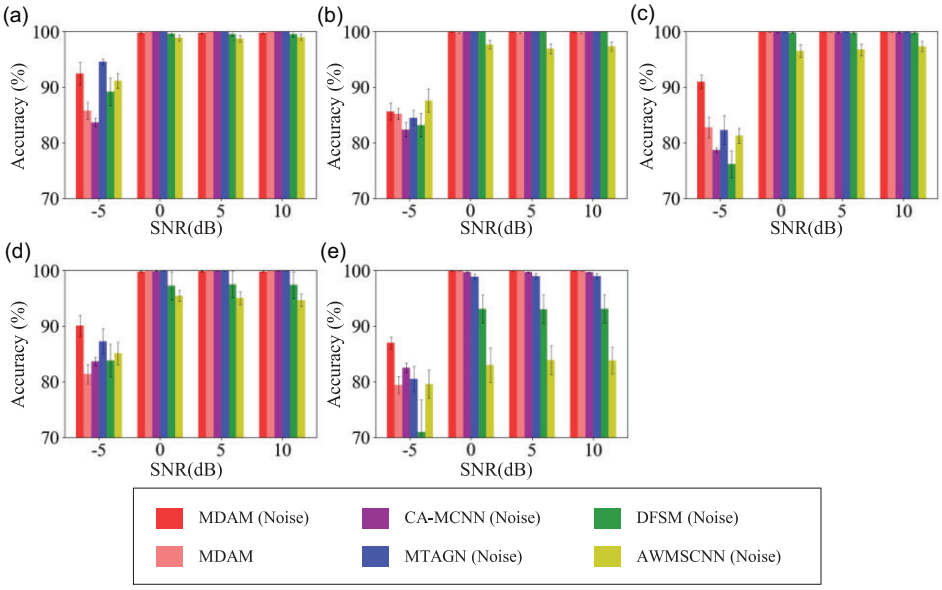

Figure 15: Accuracy results of noise-added test data for the untrained internal speed case. (a) I-1, (b) I-2, (c) I-3, (d) I-4, and (e) I-5 sub-cases.

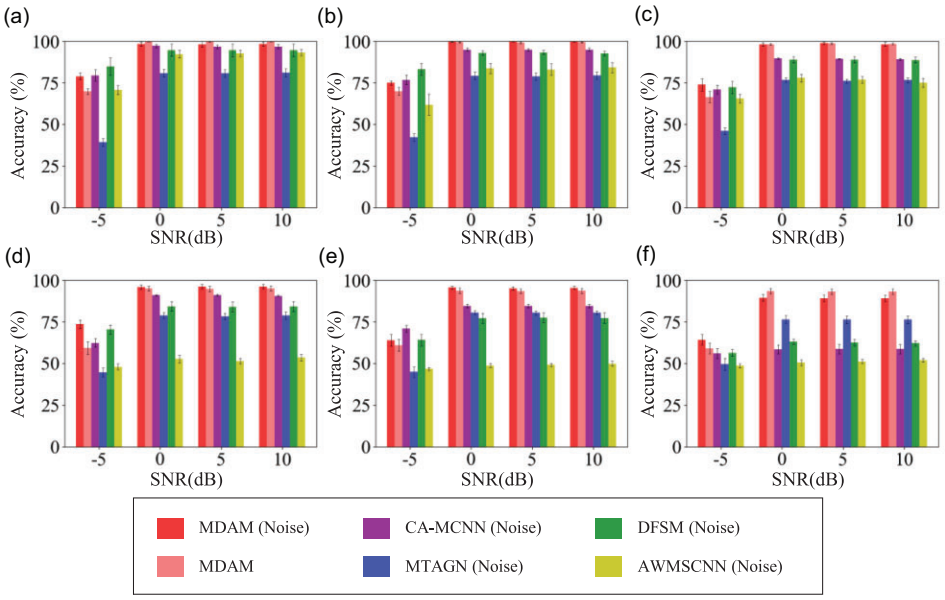

Figure 16: Accuracy results of noise-added test data for the untrained external speed case. (a) E-1, (b) E-2, (c) E-3, (d) E-4, (e) E-5, and (f) E-6 sub-cases.

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

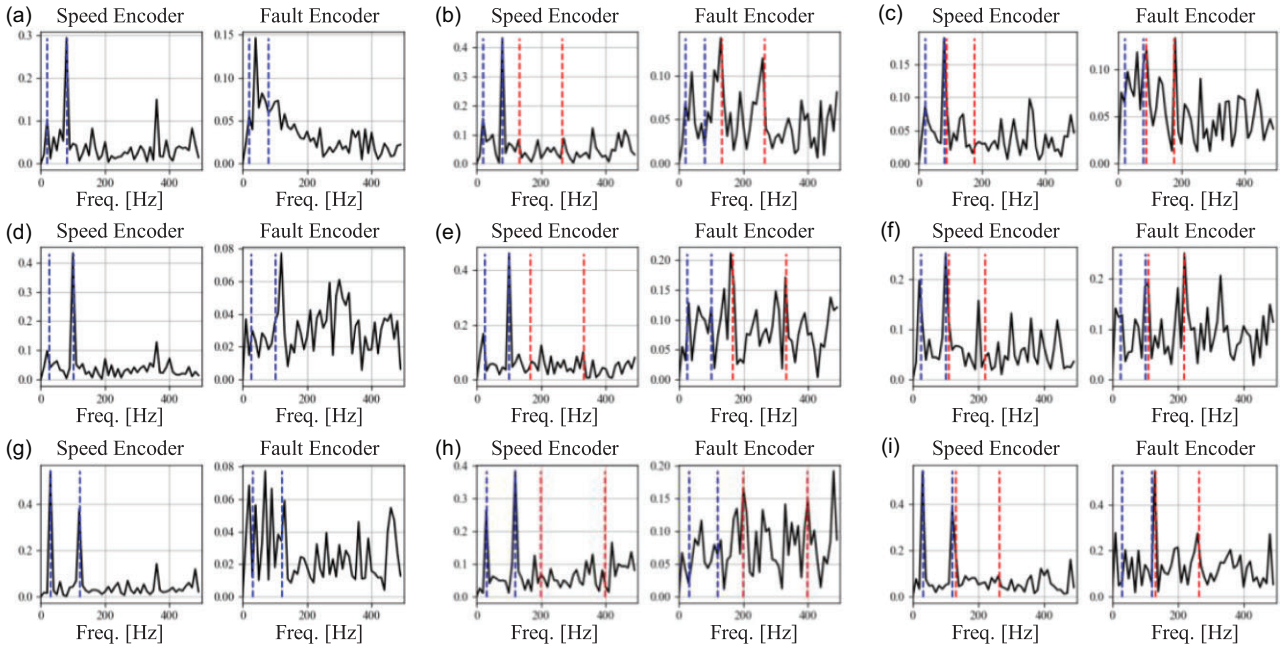

Figure 17: FFT of the 2nd hidden layer for both encoders of MDAM trained with E-6 sub-case under 1200 RPM – (a) normal, (b) inner raceway fault, and (c) outer raceway fault; 1500 RPM – (d) normal, (e) inner raceway fault, and (f) outer raceway fault; and 1800 RPM – (g) normal, (h) inner raceway fault, and (i) outer raceway fault. (Blue dotted line: rotation speed and its 4th harmonic, and red dotted line: 1st and 2nd harmonics of fault frequency).

encoder. Specifically, the features extracted from the speed encoder are expected to show the rotation speed and its harmonics rather than the fault characteristic frequencies. Conversely, the features obtained from the fault encoder are anticipated to represent the fault characteristic frequencies and not the rotation speed frequencies.

The second layer of each encoder is selected for the analysis of the extracted features, taking into consideration of the decreased Nyquist frequency and resolution caused by the pooling layer. For the hardest E-6 sub-case, fast Fourier transform (FFT) is applied to the various speed (1200, 1500, and 1800 RPM) samples for each health state. FFT results of the hidden layer feature for each encoder are represented with harmonics of rotation speed and fault characteristics for each health state in Fig. 17. Due to the physical characteristics of the testbed, the 4th harmonic of rotation frequency is also represented with FFT of hidden layer features. Throughout Fig. 17a–i, all features extracted from the speed encoder show large amplitude at rotation frequency or its 4th harmonic. In Fig. 17d, despite the low amplitude for rotation frequency, the amplitude of the 4th harmonic rotation frequency is comparatively high. At the same time, the peaks of fault frequency and its second harmonic do not appear; therefore, the speed encoder is considered to be trained to extract only speed-related information. In Fig. 17a, d, and g, features extracted by the fault encoder for normal samples did not clearly represent specific frequency peaks because of no corresponding characteristics frequency and suppressed speed-related information. Otherwise, the inner and outer raceway fault samples clearly show corresponding fault characteristics frequency rather than rotation speed frequency. In Fig. 17c, f, and i, the characteristic frequency of the outer raceway fault (4.3759∗rotation frequency) slightly overlaps with the 4th harmonic of the rotation speed, and the frequency resolution is not enough to distinguish between the two components. In detail, it can be seen that the positions of the two components are somewhat different, but there is a limit to analyzing them in this way. So, checking whether the occurrence of a second harmonic of the fault frequency is a clearer way. The features from the fault encoder clearly show 2nd harmonic components, but the speed encoder features are not depicted.

5. Conclusions

This article proposed the MDAM model for FD of REBs under various speed conditions. The proposed MH model with multi-task learning strategy enables the extraction of robust fault-related features by effectively disentangling the speed and fault-related information. For validation of the proposed method under various speed conditions, two case studies (internal and external speed conditions) were presented for test datasets. Each case study was sub-divided and deeply validated according to the sparsity of the speed conditions. The results show robust FD performance of the proposed method, even in severe test conditions. The superiority of the proposed method was also validated with a comprehensive comparison study, along with ablation and noise-parametric tests. In addition, an in-depth investigation of the feature representation through t-SNE confirmed robust performance of the proposed method. The feature analysis method t-SNE is actively applied in research using deep learning and is suitable to highlight changes in feature values according to label. Furthermore, our feature analysis using fault frequency information provides an additional discussion that MDAM is trained with physical meaning. By checking the frequency components of hidden layer features, two different types of information are extracted in separate forms.

However, there are several limitations to our research, and we plan to improve our proposed method in several ways. First, the MDAM is expected to use a general way with aspect to speed change. In this research, the speed information is dealt with a discrete way so that various speed cases can be utilized with MDAM. But usually, speed condition changes continuously, so we may re-

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

vise or improve this proposed method under continuously changing situation. The model can also be flexibly applied to different situations, such as various loads or fault health states (including different fault sizes or complex fault types). While the scope of our research is focused on various speeds, we plan to explore more improved architectures to further enhance the model's capabilities.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT; No. 2021R1A4A2001824) and National Strategic R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2021M2E6A1084687).

Conflict of interest statement

None declared.

References

Alabsi, M., Liao, Y., & Nabulsi, A. A. (2021). Bearing fault diagnosis using deep learning techniques coupled with handcrafted feature extraction: A comparative study. *Journal of Vibration and Control*, **27**, 404–414. https://doi.org/10.1177/1077546320929141.

Antoni, J., & Randall, R. B. (2003). A stochastic model for simulation and diagnostics of rolling element bearings with localized faults. *Journal of Vibration and Acoustics*, **125**, 282–289. https://doi.org/10.1115/1.1569940.

Busemeyer, J. R., Byun, E., Delosh, E. L., & McDaniel, M. A. (1997). Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. In *Studies in cognition. Knowledge, concepts and categories*(pp. 408–437). The MIT Press.

Cerrada, M., Sánchez, R. V., Li, C., Pacheco, F., Cabrera, D., Valente de Oliveira, J., & Vásquez, R. E. (2018). A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, **99**, 169–196. https://doi.org/10.1016/j.ymssp.2017.06.012.

Chen, X., Zhang, B., & Gao, D. (2021a). Bearing fault diagnosis base on multi-scale CNN and LSTM model. *Journal of Intelligent Manufacturing*, **32**, 971–987. https://doi.org/10.1007/s10845-020-01600-2.

Chen, J., Huang, R., Zhao, K., Wang, W., Liu, L., & Li, W. (2021b). Multi-scale convolutional neural network with feature alignment for bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, **70**(Icsmd 2020). https://doi.org/10.1109/TIM.2021.3077673.

Chen, P., Li, Y., Wang, K., & Zuo, M. J. (2020). A novel knowledge transfer network with fluctuating operational condition adaptation for bearing fault pattern recognition. *Measurement*, **158**, 107739. https://doi.org/10.1016/j.measurement.2020.107739.

Chen, Y., Peng, G., Xie, C., Zhang, W., Li, C., & Liu, S. (2018). ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis. *Neurocomputing*, **294**, 61–71. https://doi.org/10.1016/j.neucom.2018.03.014.

Ding, Y., Jia, M., Zhuang, J., Cao, Y., Zhao, X., & Lee, C. G. (2023). Deep imbalanced domain adaptation for transfer learning fault diagnosis of bearings under multiple working conditions. *Reliability Engineering and System Safety*, **230**, 108890. https://doi.org/10.1016/j.ress.2022.108890.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. http://www.deeplearningbook.org.

Guo, W., & Tse, P. W. (2013). A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals. *Journal of Sound and Vibration*, **332**, 423–441. https://doi.org/10.1016/j.jsv.2012.08.017.

Hasan, M. J., Islam, M. M. M., & Kim, J. M. (2019). Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions. *Measurement*, **138**, 620–631. https://doi.org/10.1016/j.measurement.2019.02.075.

Huang, Y. J., Liao, A. H., Hu, D. Y., Shi, W., & Zheng, S. B. (2022). Multi-scale convolutional network with channel attention mechanism for rolling bearing fault diagnosis. *Measurement*, **203**, 111935. https://doi.org/10.1016/J.MEASUREMENT.2022.111935.

Jang, J.-G., Noh, C.-M., Kim, S.-S., Shin, S.-C., Lee, S.-S., & Lee, J.-C. (2023). Vibration data feature extraction and deep learning-based preprocessing method for highly accurate motor fault diagnosis. *Journal of Computational Design and Engineering*, **10**, 204–220. https://doi.org/10.1093/jcde/qwac128.

Jiang, Q., Chang, F., & Sheng, B. (2019). Bearing fault classification based on convolutional neural network in noise environment. *IEEE Access*, **7**, 69795–69807. https://doi.org/10.1109/ACCESS.2019.2919126.

Jin, X., Zhao, M., Chow, T. W. S., & Pecht, M. (2014). Motor bearing fault diagnosis using trace ratio linear discriminant analysis. *IEEE Transactions on Industrial Electronics*, **61**, 2441–2451. https://doi.org/10.1109/TIE.2013.2273471.

Kim, H., Park, C. H., Suh, C., Chae, M., Yoon, H., & Youn, B. D. (2023b). MPARN: Multi-scale path attention residual network for fault diagnosis of rotating machines. *Journal of Computational Design and Engineering*, **10**, 860–872. https://doi.org/10.1093/jcde/qwad031.

Kim, H., & Youn, B. D. (2019). A new parameter repurposing method for parameter transfer with small dataset and its application in fault diagnosis of rolling element bearings. *IEEE Access*, **7**, 46917–46930. https://doi.org/10.1109/ACCESS.2019.2906273.

Kim, K., Yoon, H., & Youn, B. D. (2023a). A noise–robust feature extraction method for rolling element bearing diagnosis: Linear power–normalized cepstral coefficients (LPNCC). *International Journal of Precision Engineering and Manufacturing-Green Technology*, **10**, 217–232. https://doi.org/10.1007/s40684-022-00448-6.

Kim, S. J., Kim, K., Hwang, T., Park, J., Jeong, H., Kim, T., & Youn, B. D. (2022). Motor-current-based electromagnetic interference de-noising method for rolling element bearing diagnosis using acoustic emission sensors. *Measurement*, **193**, 110912. https://doi.org/10.1016/j.measurement.2022.110912.

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 2015 3rd International Conference on Learning Representations (ICLR)*(pp. 1–15). ICLR 2015.

Li, S., An, Z., & Lu, J. (2020). A novel data-driven fault feature separation method and its application on intelligent fault diagnosis under variable working conditions. *IEEE Access*, **8**, 113702–113712. https://doi.org/10.1109/ACCESS.2020.2996713.

Lin, J., Shao, H., Min, Z., Luo, J., Xiao, Y., Yan, S., & Zhou, J. (2022). Cross-domain fault diagnosis of bearing using improved semi-supervised meta-learning towards interference of out-of-distribution samples. *Knowledge-Based Systems*, **252**, 109493. https://doi.org/10.1016/J.KNOSYS.2022.109493.

Lin, M., Chen, Q., & Yan, S. (2014). Network in network. In *Proceedings of the 2014 2nd International Conference on Learning Representations (ICLR)*(pp. 1–10). ICLR 2014.

Liu, Z., Tang, X., Wang, X., Mugica, J. E., & Zhang, L. (2021). Wind turbine blade bearing fault diagnosis under fluctuating speed operations via Bayesian augmented Lagrangian analysis. *IEEE Transactions on Industrial Informatics*, **17**, 4613–4623. https://doi.org/10.1109/TII.2020.3012408.

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

Liu, Z., Wang, H., Liu, J., Qin, Y., & Peng, D. (2020). Multi-task learning based on lightweight 1DCNN for fault diagnosis of wheelset bearings. *IEEE Transactions on Instrumentation and Measurement*, **9456**, 1–1. https://doi.org/10.1109/tim.2020.3017900.

Lu, C., Wang, Z. Y., Qin, W. L., & Ma, J. (2017a). Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing*, **130**, 377–388. https://doi.org/10.1016/j.sigpro.2016.07.028.

Nayana, B. R., & Geethanjali, P. (2017). Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal. *IEEE Sensors Journal*, **17**, 5618–5625. https://doi.org/10.1109/JSEN.2017.2727638.

Nikula, R. P., Karioja, K., Pylvänäinen, M., & Leiviskä, K. (2020). Automation of low-speed bearing fault diagnosis based on autocorrelation of time domain features. *Mechanical Systems and Signal Processing*, **138**, 106572. https://doi.org/10.1016/j.ymssp.2019.106572.

Oh, H., Lee, Y., Lee, J., Joo, C., & Lee, C. (2022). Feature selection algorithm based on density and distance for fault diagnosis applied to a roll-to-roll manufacturing system. *Journal of Computational Design and Engineering*, **9**, 805–825. https://doi.org/10.1093/jcde/qwac028.

Peng, D., Wang, H., Liu, Z., Zhang, W., Zuo, M. J., & Chen, J. (2020). Multibranch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition. *IEEE Transactions on Industrial Informatics*, **16**, 4949–4960. https://doi.org/10.1109/TII.2020.2967557.

Qiao, H., Wang, T., Wang, P., Zhang, L., & Xu, M. (2019). An adaptive weighted multiscale convolutional neural network for rotating machinery fault diagnosis under variable operating conditions. *IEEE Access*, **7**, 118954–118964. https://doi.org/10.1109/ACCESS.2019.2936625.

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics-A tutorial. *Mechanical Systems and Signal Processing*, **25**, 485–520. https://doi.org/10.1016/j.ymssp.2010.07.017.

Raouf, I., Lee, H., & Kim, H. S. (2022). Mechanical fault detection based on machine learning for robotic RV reducer using electrical current signature analysis: A data-driven approach. *Journal of Computational Design and Engineering*, **9**, 417–433. https://doi.org/10.1093/jcde/qwac015.

Rauber, T. W., De Assis Boldt, F., & Varejão, F. M. (2015). Heterogeneous feature models and feature selection applied to bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, **62**, 637–646. https://doi.org/10.1109/TIE.2014.2327589.

Sharma, V., & McNeill, J. H. (2009). To scale or not to scale: The principles of dose extrapolation. *British Journal of Pharmacology*, **157**, 907–921. https://doi.org/10.1111/j.1476-5381.2009.00267.x.

Shi, Z., Chen, J., Zi, Y., & Zhou, Z. (2021). A novel multitask adversarial network via redundant lifting for multicomponent intelligent fault detection under sharp speed variation. *IEEE Transactions on Instrumentation and Measurement*, **70**, 1–10. https://doi.org/10.1109/TIM.2021.3055821.

Ternes, L., Dane, M., Gross, S., Labrie, M., Mills, G., Gray, J., & Chang, Y. H. (2022). A multi-encoder variational autoencoder controls multiple transformational features in single-cell image analysis. *Communications Biology*, **5**, 1–10. https://doi.org/10.1038/s42003-022-03218-x.

Tyagi, S., & Panigrahi, S. K. (2017). An improved envelope detection method using particle swarm optimisation for rolling element bearing fault diagnosis. *Journal of Computational Design and Engineering*, **4**, 305–317. https://doi.org/10.1016/j.jcde.2017.05.002.

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, **9**, 2579–2605.

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*(pp. 1096–1103). ICML 2008. https://doi.org/10.1145/1390156.1390294.

Wang, H., Liu, Z., Peng, D., & Qin, Y. (2019). Understanding and learning discriminant features based on multi-attention 1DCNN for wheelset bearing fault diagnosis. *IEEE Transactions on Industrial Informatics*, **16**, 1–1. https://doi.org/10.1109/tii.2019.2955540.

Wang, H., Liu, Z., Peng, D., Yang, M., & Qin, Y. (2022). Feature-level attention-guided multitask CNN for fault diagnosis and working conditions identification of rolling bearing. *IEEE Transactions on Neural Networks and Learning Systems*, **33**, 4757–4769. https://doi.org/10.1109/TNNLS.2021.3060494.

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**, 107327. https://doi.org/10.1016/j.ymssp.2020.107327.

Xie, Z., Chen, J., Feng, Y., Zhang, K., & Zhou, Z. (2022). End to end multitask learning with attention for multi-objective fault diagnosis under small sample. *Journal of Manufacturing Systems*, **62**, 301–316. https://doi.org/10.1016/j.jmsy.2021.12.003.

Yao, J., & Han, T. (2023). Data-driven lithium-ion batteries capacity estimation based on deep transfer learning using partial segment of charging/discharging data. *Energy*, **271**, 127033. https://doi.org/10.1016/J.ENERGY.2023.127033.

Yu, X., Wang, Y., Liang, Z., Shao, H., Yu, K., & Yu, W. (2023). An adaptive domain adaptation method for rolling bearings fault diagnosis fusing deep convolution and self-attention networks. *IEEE Transactions on Instrumentation and Measurement*, **72**, 1–1. https://doi.org/10.1109/tim.2023.3246494.

Zhao, B., Zhang, X., Zhan, Z., & Wu, Q. (2021). Deep multi-scale separable convolutional network with triple attention mechanism: A novel multi-task domain adaptation method for intelligent fault diagnosis. *Expert Systems with Applications*, **182**, 115087. https://doi.org/10.1016/j.eswa.2021.115087.

Downloaded from https://academic.oup.com/jcde/article/10/4/1804/7234338 by guest on 15 December 2025

Received: February 23, 2023. **Revised:** July 17, 2023. **Accepted:** July 17, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of the Society for Computational Design and Engineering. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com