Multi-source Variable Decoupling Network for Compound Fault Diagnosis of Train Bogie

Qitao Yin[1], Zhibin Guo[2], Tiantian Wang[1,2], Jingsong Xie[2*] and Jinsong Yang[2]

[1] *School of Mechanical and Vehicle Engineering, Hunan University, Changsha, Hunan, 410082, China*
yin_qitao@163.com
wangtt@hnu.edu.cn

[2] *School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, 410075, China*
marcogzb@csu.edu.cn
jingsongxie@foxmail.com
yangjs@csu.edu.cn

ABSTRACT

When compound faults occur in rotating machinery, the mutual coupling and interference among different fault sources make it extremely difficult to directly isolate individual faults from the observed signals. Therefore, fault decoupling is essential prior to diagnosis. In this study, we propose a semi-supervised multi-source variable decoupling network (MVD-Net) that enables blind separation of unknown compound fault signals using only single-fault samples for training. First, low-dimensional features are extracted from the mixed signal through an encoder. These features are then mapped to multiple independent latent spaces corresponding to different fault sources via variational inference, while the number of sources is adaptively estimated using the evidence lower bound (ELBO). Subsequently, each source-specific decoder generates an estimated source signal from its corresponding latent representation. To ensure that each decoder focuses on a distinct fault component, a source-selective activation mechanism is incorporated into the decoding process, effectively mitigating the random assignment issue commonly encountered in traditional blind source separation methods. Finally, based on the estimated source signals, a separation mask is derived to extract individual sources from the original mixed signal. Two compound fault decoupling and diagnosis experiments were conducted on the BJTU-RAO dataset. The results demonstrate that compared with other methods, the proposed approach yields cleaner separated signals with more distinct time-frequency fault features and achieves higher diagnostic accuracy.

1. INTRODUCTION

The train bogie contains a large number of rotating components that play critical roles in power transmission and load-bearing. Due to long-term service under high-speed and heavy-load conditions, combined with factors such as material aging and poor lubrication, key rotating parts such as gears and bearings are prone to various faults（He, Wu, Jin, Huang, Wei, & Yi, 2025). However, early weak faults often do not trigger immediate shutdowns for maintenance, making them likely to induce and accumulate additional faults, thereby leading to compound faults. The most prominent characteristic of compound faults is the coupling and mutual interference between different fault features（Xiao & Yu, 2023). Even with noise reduction processing, it remains challenging to effectively extract fault characteristics, posing significant difficulties for bogie monitoring and maintenance.

For the problem of compound fault diagnosis, early attempts involved treating each compound fault mode as an independent fault type for diagnosis. For example, Zhong, Yang, and Wong (2010) proposed a support vector machine (SVM)-based fault diagnosis method capable of identifying compound faults through feature extraction. While such methods perform well for known compound fault patterns, they suffer from a significant drawback: in industrial settings, obtaining training samples for all possible compound fault modes is extremely difficult. Consequently, research has shifted focus toward decoupling compound faults of unknown patterns into a limited set of single faults before diagnosis.

Classical signal decomposition methods include wavelet decomposition, empirical mode decomposition (EMD), variational mode decomposition (VMD), and deconvolution. Jiang, Zhu, and Li (2016) proposed a compound faults

detection method for rolling bearings based on empirical wavelet transform and chaotic oscillator in Chaos. For compound fault decoupling, multi-wavelet and EMD methods were employed. Wan, Zhang, and Dou (2018) developed an improved fast spectral kurtosis with VMD to identify weak fault features in compound bearing faults. Lyu, Hu, Zhou, and Wang (2019) introduced a quantum genetic algorithm-enhanced maximum correlated kurtosis deconvolution (QGA-MCKD) method applicable to gear and bearing compound fault diagnosis. Zhang, Li, Xin, and Ma (2020) proposed an intrinsic component filtering-based approach for compound fault diagnosis that could adaptively utilize different filters. However, these classical signal processing-based compound fault diagnosis methods typically require substantial expert experience and prior knowledge of fault mechanisms to properly set parameters in order to achieve satisfactory signal decoupling quality and diagnostic accuracy.

In recent years, with the continuous development of artificial intelligence technology, intelligent decoupling diagnosis methods for compound faults based on deep learning have begun to emerge. Huang, Liao, Zhang, and Li (2018) proposed a Deep Decoupling Convolutional Neural Network (DDCNN), designing multi-layer capsules as decoupling classifiers to accurately identify and decouple compound faults. Jin, Qin, Huang, and Liu (2021) introduced a novel Decoupling Attention Residual Network (DARN) for compound fault diagnosis, which utilizes attention modules and multi-label decoupling classifiers to precisely decouple and identify compound faults. He, Chu, Li, Zhang, Wei, and Hu (2023) developed a multi-fault coupling photovoltaic system diagnosis scheme based on multi-label classification, decoupling compound faults into multiple single faults through a multi-label learning approach. However, these intelligent decoupling methods typically perform decoupling at either deep feature or semantic levels, resulting in poor interpretability. Additionally, many methods still rely on comprehensive labeled compound fault samples to achieve relatively ideal diagnostic accuracy.

In contrast, unsupervised blind source separation (BSS) methods can separate multi-source mixed signals into independent source signals without prior knowledge of signal sources or mixing systems. When applied to compound fault diagnosis, BSS requires neither prior fault knowledge nor expert experience, and is independent of compound fault samples. Zhang, Ji, Huang, and Lou (2021) proposed a canonical correlation analysis (CCA)-based method for blind extraction of specific fault signals from multi-channel observations, applied to compound fault diagnosis in aero-engine main shaft bearings. Peng, Zhu, Teng, and Liu (2023) developed a Generalized Gaussian cyclostationary (GGCS) model based on fault frequency synchronous averaging, effectively incorporating shape parameter estimation into blind deconvolution methods. Cai and Tang (2024) introduced a novel blind deconvolution method called spectral sparse entropy ratio blind deconvolution (SSERBD) that requires no prior fault knowledge and effectively attenuates transmission path effects. Dong, Feng, Liu, and Zhang (2025) improved the maximum second-order cyclostationarity blind deconvolution (CYCBD) method by proposing adaptive blind deconvolution (ABDD), which outperforms traditional methods in compound fault feature extraction. However, BSS faces challenges in determining the number and order of source signals. Additionally, the inability to pre-learn source signal information limits its effectiveness in separating compound fault signals.

To address the aforementioned challenges, we propose a semi-supervised Multi-source Variational Decoupling Network (MVD-Net). First, an encoder extracts low-dimensional features from compound fault signals. These features are then mapped to multiple independent latent spaces corresponding to different fault sources through variational inference. Subsequently, independent decoders generate source estimates from each separated channel's latent representation. Finally, separation masks are computed based on source estimates, which are applied to the original mixed signals to obtain separated signals. The main contributions of this study are as follows:

1. We propose a semi-supervised Multi-source Variational Decoupling Network that requires only single-fault samples for training. This achieves data-driven semi-blind source separation, effectively utilizing single-fault sample information while eliminating dependence on compound fault samples.
2. Under the evidence lower bound (ELBO) of variational inference, MVD-Net can automatically determine the number of fault sources.
3. A source-selective activation mechanism is incorporated in the decoding process, ensuring each separation channel focuses on one fault type. This effectively mitigates the random semantic assignment problem caused by BSS's inherent ordering uncertainty.

The remainder of this paper is organized as follows. Section 2 introduces the fundamental theory of BSS and analyzes its two types of uncertainty, along with the basic principles of variational autoencoders. Section 3 details the proposed methodology's overall framework and implementation. Section 4 presents two compound fault decoupling and diagnosis experiments using the BJTU-RAO dataset. Section 5 provides discussion and conclusions.

2. FUNDAMENTAL CONCEPTS AND THEORIES

2.1. The uncertainty of BSS

BSS is a method for recovering source signals from observed mixtures without prior knowledge of either the source signals or the mixing system. Consider N statistically independent

signal vectors $\boldsymbol{x}_1(t)$, $\boldsymbol{x}_2(t)$, ..., $\boldsymbol{x}_n(t)$ from source signals that are mixed through an $M \times N$ mixing matrix $\boldsymbol{A}$, yielding M observed signal $\boldsymbol{s}_1(t)$, $\boldsymbol{s}_2(t)$, ..., $\boldsymbol{s}_m(t)$. The signal model is expressed as:

$$\boldsymbol{s}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{n}(t) \tag{1}$$

where $\boldsymbol{n}(t)$ represents an $M \times 1$ noise vector.

The BSS problem can therefore be formulated as estimating an $N \times M$ separation matrix $\boldsymbol{W}$ (the left inverse of $\boldsymbol{A}$, satisfying $\boldsymbol{W}\boldsymbol{A} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix), which recovers the source signals through:

$$\boldsymbol{y}(t) = \boldsymbol{W}\boldsymbol{s}(t) = \boldsymbol{W}\boldsymbol{A}\boldsymbol{x}(t) \approx \boldsymbol{x}(t) \tag{2}$$

Since both the mixing matrix $\boldsymbol{A}$ and source signals $\boldsymbol{x}(t)$ in Eq (1) are unknown, BSS suffers from two inherent uncertainties:

1. Source Number Uncertainty: The dimensionality of separation matrix W cannot be determined a priori, typically requiring an assumed maximum number of potential sources.
2. Permutation Ambiguity: The correspondence between separated signals $\boldsymbol{y}(t)$ and original source components $\boldsymbol{x}(t)$ remains indeterminate.

These uncertainties pose critical challenges for fault diagnosis: Permutation ambiguity causes semantic inconsistency in separated channels, requiring all fault types to be evaluated for every channel, thereby reducing diagnostic efficiency and accuracy. Source number uncertainty may lead directly to either missed detections or false alarms (overestimation).

2.2. Variational Autoencoder

The Variational Autoencoder (VAE) is a generative model composed of an encoder, latent variables, and a decoder. Its core concept revolves around data "compression and reconstruction". For a given set of data samples $\boldsymbol{X} = \boldsymbol{x}_1(t), \boldsymbol{x}_2(t), ..., \boldsymbol{x}_m(t)$, if we can determine the probability distribution $p(x)$, we can generate all possible approximate samples. However, directly computing $p(\boldsymbol{x})$ is extremely challenging. Instead, $p(\boldsymbol{x})$ is assumed to be generated from latent variables $\boldsymbol{z}$ that follow a Gaussian distribution:

$$p(\boldsymbol{x}) = \sum_{z} p(\boldsymbol{x} \mid \boldsymbol{z}) p(\boldsymbol{z}) \tag{3}$$

The corresponding decoder is defined as:

$$p(\boldsymbol{x} \mid \boldsymbol{z}) = N\left(\boldsymbol{x};\ \mu(\boldsymbol{z}),\ \sigma^2(\boldsymbol{z})\mathbf{I}\right) \tag{4}$$

Where μ represents the mean and σ^2 the variance. Accordingly, the encoder's task is to compress input data into a probability distribution:

$$q(\boldsymbol{z} \mid \boldsymbol{x}) = N\left(\boldsymbol{z};\ \mu(\boldsymbol{x}),\ \sigma^2(\boldsymbol{x})\mathbf{I}\right) \tag{5}$$

Since computing $p(\boldsymbol{x})$ is equivalent to computing $\log p(\boldsymbol{x})$, and given that $\sum_{z} q(\boldsymbol{z} \mid \boldsymbol{x}) = 1$, we derive:

$$\begin{aligned} \log(p(\boldsymbol{x})) &= \sum_{z} q(z \mid \boldsymbol{x}) \log(p(\boldsymbol{x})) \\ &= \sum_{z} q(z \mid \boldsymbol{x}) \log\left(\frac{p(z,\boldsymbol{x})}{q(z \mid \boldsymbol{x})} \frac{q(z \mid \boldsymbol{x})}{p(z \mid \boldsymbol{x})}\right) \\ &= \sum_{z} q(z \mid \boldsymbol{x}) \log\left(\frac{p(z,\boldsymbol{x})}{q(z \mid \boldsymbol{x})}\right) + \sum_{z} q(z \mid \boldsymbol{x}) \log\left(\frac{q(z \mid \boldsymbol{x})}{p(z \mid \boldsymbol{x})}\right) \end{aligned} \tag{6}$$

Here, the second term represents the Kullback-Leibler (KL) divergence between two distributions and is non-negative. By denoting the first term as L_b, we obtain:

$$\log(p(\boldsymbol{x})) \geq L_b \tag{7}$$

Thus, we arrive at the lower bound of $\log(p(\boldsymbol{x}))$, known as the ELBO. Maximizing $p(\boldsymbol{x})$ is equivalent to maximizing L_b.

3. PROPOSED METHOD

3.1. Overall Framework of the Proposed Method

In current research on compound fault signal decoupling, unsupervised BSS-based methods suffer from two critical limitations: (1) difficulty in determining the number and order of source signals, and (2) inability to leverage prior knowledge of source signal characteristics. On the other hand, data-driven intelligent decoupling methods still rely heavily on labeled compound fault samples and exhibit poor interpretability.

To address these challenges, we propose a semi-supervised Multi-source Variational Decoupling Network (MVD-Net). MVD-Net not only resolves the inherent uncertainties of traditional BSS methods (i.e., source number and permutation ambiguity) to achieve blind separation of compound fault signals but also incorporates prior distribution features learned from single-fault samples. This hybrid approach combines the strengths of both unsupervised decoupling and data-driven methods. The overall framework and forward computation process of MVD-Net are illustrated in Figure 1.

Step1: Assuming a maximum of K potential sources, the encoder extracts feature $\boldsymbol{h}$ via Eq. (5) and projects them into K independent latent variables $(\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k^2)$ (k=1, 2, …, K) through dedicated fully-connected (FC) layers:

$$\begin{aligned} \boldsymbol{\mu} &= \boldsymbol{W}_{\mu}\boldsymbol{h} + \boldsymbol{b}_{\mu} \\ \log\boldsymbol{\sigma}^2 &= \boldsymbol{W}_{log\sigma^2}\boldsymbol{h} + \boldsymbol{b}_{log\sigma^2} \end{aligned} \tag{8}$$

where $\boldsymbol{w}$, $\boldsymbol{b}$ denote trainable parameters for the FC. This achieves low-dimensional feature decoupling in latent space.

Step2: To enable backpropagation, latent variables are sampled using the reparameterization trick:

$$\boldsymbol{z}_k = \boldsymbol{\mu}_k + \boldsymbol{\sigma}_k \odot \boldsymbol{\varepsilon} \tag{9}$$

Where $\odot$ denotes element-wise multiplication, $\boldsymbol{\varepsilon} \sim N(0,1)$. This maintains differentiability while preserving stochasticity.

Step3: Each latent variable $\boldsymbol{z}_k$ is fed into an independent decoder to reconstruct the k-th source estimate $\hat{\boldsymbol{x}}_k$.

Step4: Separation masks $\boldsymbol{m}_k$ are computed from source estimates $\hat{\boldsymbol{x}}_k$. Applying masks to the raw mixed signal $\boldsymbol{x}$ yields final separated signals $\tilde{\boldsymbol{x}}_k$.

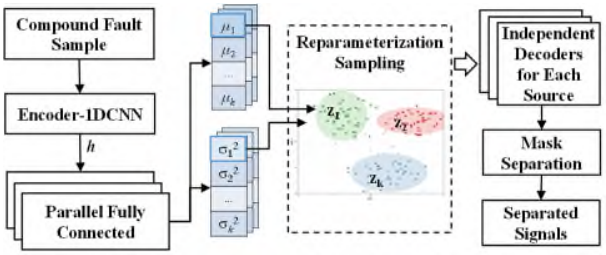


Figure 1. The overall framework and forward computation process of MVD-Net

3.2. Optimization Objective

Based on Eq (5), the ELBO can be further decomposed as:

$$L_b = \sum_z q(\boldsymbol{z}|\ \boldsymbol{x}) \log\left(\frac{p(\boldsymbol{z})}{q(\boldsymbol{z}|\ \boldsymbol{x})}\right) + \sum_z q(\boldsymbol{z}|\ \boldsymbol{x}) \log(p(\boldsymbol{x}|\ \boldsymbol{z})) \tag{10}$$

According to the definitions of KL divergence and expectation, the ELBO can ultimately be expressed as:

$$L_b = -D_{KL}(q(\boldsymbol{z}|\ \boldsymbol{x}) \| p(\boldsymbol{z})) + \mathrm{E}_{q(\boldsymbol{z}|\ \boldsymbol{x})}(\log(p(\boldsymbol{x}|\ \boldsymbol{z}))) \tag{11}$$

By taking the negative of the ELBO, we obtain the loss function for the VAE:

$$L_{\mathrm{VAE}} = D_{KL}(q(\boldsymbol{z}\mid \boldsymbol{x}) \| p(\boldsymbol{z})) - \mathrm{E}_{q(\boldsymbol{z}|\boldsymbol{x})}(\log(p(\boldsymbol{x}\mid \boldsymbol{z}))) \tag{12}$$

Here, the first term on the right-hand side is the KL divergence loss, which encourages the distribution of the encoder's output $\boldsymbol{z}$ to approximate a standard normal distribution, providing regularization. The second term is the reconstruction loss, which ensures the generated data closely matches the original data.

The primary structure of MVD-Net can be viewed as a multi-channel extension of the VAE. Therefore, its loss function also includes KL divergence loss and reconstruction loss, albeit with slightly different interpretations and computations.

According to Eq (9) and the definition of KL divergence, the KL divergence loss for the K-channel MVD-Net is defined as the average of the KL divergences across all source channels:

$$\begin{aligned} L_{\mathrm{MVD}}^{KL} &= \frac{1}{K}\sum_{k=1}^{K} D_{KL}(q(\boldsymbol{z}\mid \boldsymbol{x}) \| p(\boldsymbol{z})) \\ &= \frac{1}{K}\sum_{k=1}^{K}\left[-\frac{1}{2}\sum_{j=1}^{D_z}(1+\log\sigma_{k,j}^2 - \mu_{k,j}^2 - \sigma_{k,j}^2)\right] \end{aligned} \tag{13}$$

where D_z is the dimensionality of the latent variables.

MVD-Net further requires the fusion of all source estimates $\hat{\boldsymbol{x}}_k$ generated by the decoders to obtain the reconstructed signal $\hat{\boldsymbol{x}}$:

$$\hat{\boldsymbol{x}} = \sum_{k=1}^{K} w_k \hat{\boldsymbol{x}}_k \tag{14}$$

where w represents the channel activation weights, the specific role of which will be discussed in the next section.

Since the reconstruction loss measures the discrepancy between the reconstructed signal and the original signal, the mean squared error (MSE) loss is used as the reconstruction loss for MVD-Net:

$$\boldsymbol{L}_{\mathrm{MVD}}^{recon} = \| \hat{\boldsymbol{x}} - \boldsymbol{x} \|^2 \tag{15}$$

The total loss for MVD-Net is:

$$\boldsymbol{L}_{\mathrm{MVD}} = \alpha \boldsymbol{L}_{\mathrm{MVD}}^{KL} + \beta \boldsymbol{L}_{\mathrm{MVD}}^{recon} \tag{16}$$

where α, β are weighting coefficients.

The role of $\boldsymbol{L}_{\mathrm{MVD}}^{KL}$ is to drive the separated signals of each channel toward Gaussian noise, while $\boldsymbol{L}_{\mathrm{MVD}}^{recon}$ ensures the separated signals approximate single-fault signals. Under the balance of these two losses, the model can learn distinct fault features while avoiding over-separation into non-existent faults, thereby achieving automatic source number identification.

3.3. Source-Selective Activation Training Mechanism

MVD-Net requires training with both single-fault samples and mixed samples. The mixed samples are simulated compound fault samples randomly synthesized from different types of single-fault samples, without the need to know their specific fault types. Single-fault samples are labeled according to their fault type as i ($i = 0, 1, ..., K$), while mixed samples are uniformly labeled as −1. Each training

batch contains both single-fault and mixed samples, and the training process is illustrated in Figure 2.

First, the encoder extracts low-dimensional features from the input samples. These features are then decoupled into *K* latent variables through parallel fully-connected layers using multi-source variational inference, and the KL divergence loss is computed. Next, the multi-source latent variables undergo reparameterization sampling and are fed into independent decoders for signal decoding and reconstruction, during which the reconstruction loss is calculated.

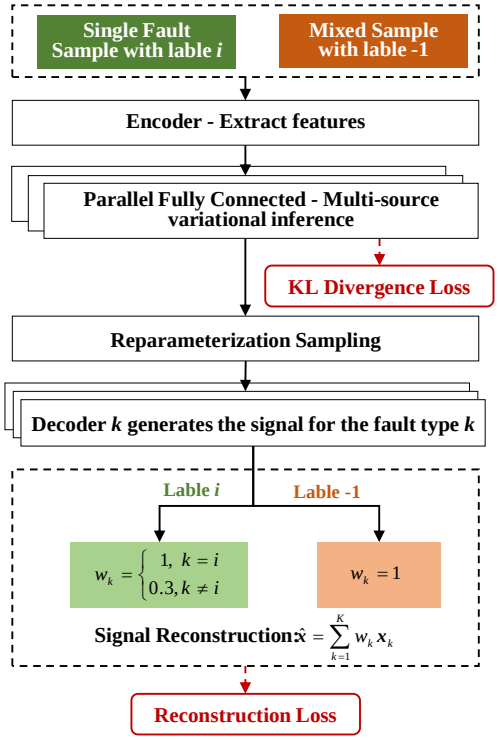


Figure 2. The training process of MVD-Net

During the signal decoding and reconstruction phase, each decoder is associated with a specific fault type, serving as a specialized separation channel for that fault, denoted as decoder-k. It is important to note that this association between channels and fault types is a soft binding implemented through the source-selective activation mechanism, rather than a hard-coded strong binding. This allows the model to learn channel assignments autonomously, promoting better feature learning. The principles of the source-selective activation mechanism are as follows:

1. When the latent variables of a single-fault sample are input to the decoder group, the signal generated by the decoder-k bound to that fault type is assigned a high weight during signal reconstruction. Signals generated by other decoders are assigned low weights during reconstruction.
2. When the latent variables of a mixed sample are input to the decoders, all signals generated by the decoders are assigned equal weights during reconstruction.

This mechanism serves to specialize each separation channel for a specific fault type, which not only determines the order of separated signals but also fixes the fault semantics of each channel, thereby improving diagnostic efficiency and accuracy. Furthermore, this mechanism constitutes a key design element that enables the trainability of the BSS model. It allows the model to simultaneously learn distinct fault patterns from single-fault samples while acquiring the capability to decouple mixed signals from composite samples.

3.4. Mask-Based Separation

To precisely capture fine-grained details from the original signal, a post-training masking operation is applied to convert source estimates $\hat{\boldsymbol{x}}_k$ from MVD-Net into separated signals $\tilde{\boldsymbol{x}}_k$:

$$\tilde{\boldsymbol{x}}_k = \boldsymbol{m}_k \cdot \boldsymbol{x} \tag{17}$$

$$\boldsymbol{m}_k = \frac{|\hat{\boldsymbol{x}}_k|}{\sum_{j=1}^{K} |\hat{\boldsymbol{x}}_j| + \grave{o}} \tag{18}$$

where ϵ is an infinitesimal constant to prevent division by zero.

Mask separation converts the source estimates generated by the decoder into weight masks to precisely extract each source signal from the original signal. This method preserves all frequency details and characteristics of the original signal while ensuring energy conservation (the sum of separated signals strictly equals the original signal). It avoids the potential detail loss and energy mismatch issues that may arise from directly using source estimate signals, achieving an optimal combination of automatic source identification and precise signal separation.

4. EXPERIMENTS AND VERIFICATION

4.1. Dataset

The experiments were validated using the BJTU dataset. As shown in Figure 3, the test bench was built at a 1:2 scale based on an actual metro train bogie, containing fault data from both the gearbox and axle box components. The fault data was collected by vibration acceleration sensors with a sampling frequency of 64 kHz.

The gearbox includes bearing inner race faults (IF), gear tooth wear faults (WT), and their compound faults (IR&WT). The axle box includes bearing outer race faults (OF), bearing roller faults (BF), and their compound faults (OR&BF).

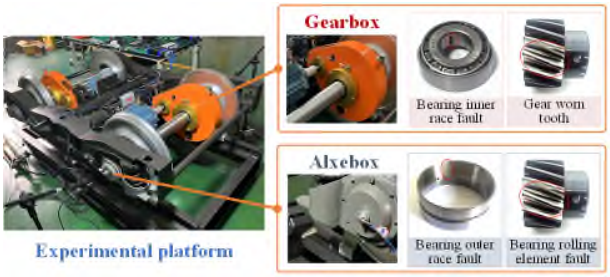
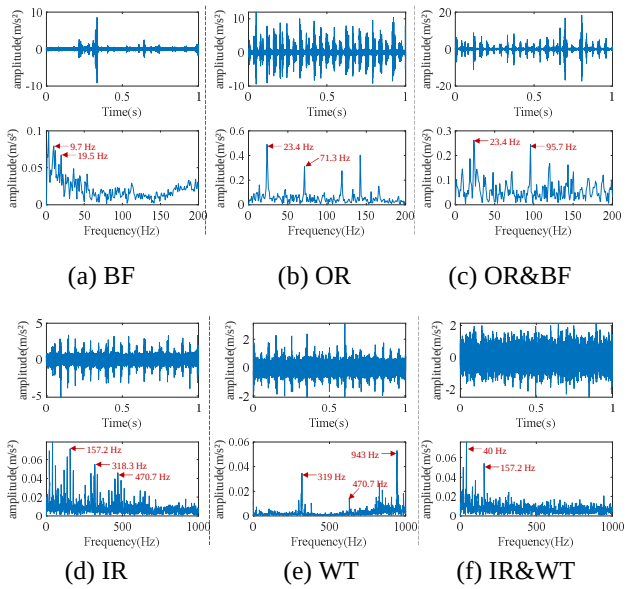


Figure 3. The test bench and faulty parts

The time-domain waveforms and frequency spectra (or envelope spectra) of all samples are shown in Figure 4. Among them, gear faults are represented by Fourier spectra, while other signals are represented by envelope spectra. For single-fault signals, characteristic frequencies and harmonics are clear in frequency or envelope spectra, while compound fault features are obscured.



(a) BF (b) OR (c) OR&BF

(d) IR (e) WT (f) IR&WT

Figure 4. The time-domain waveforms and frequency spectra (or envelope spectra) of all samples

The compound fault decoupling experiments for the two components are labeled Task A and Task B. Test conditions use a 20Hz motor speed and 10kN load. Dataset details are in Table 1: both tasks follow the same partitioning—300 single-fault samples and 900 mixed samples for training, and 300 unlabeled compound fault samples for testing.

Table 1. The dataset information.

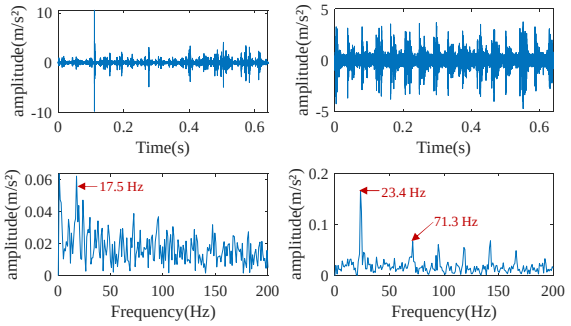
Task	Dataset Split	Fault type	Sample size	Label	Batch Ratio
A	Training	OR	300	0	20%
		BF	300	0	20%
		Mixed	900	-1	60%
	Testing	OR&BF	300	—	—
B	Training	IR	300	0	20%
		WT	300	0	20%
		Mixed	900	-1	60%
	Testing	IR&WT	300	—	—

.

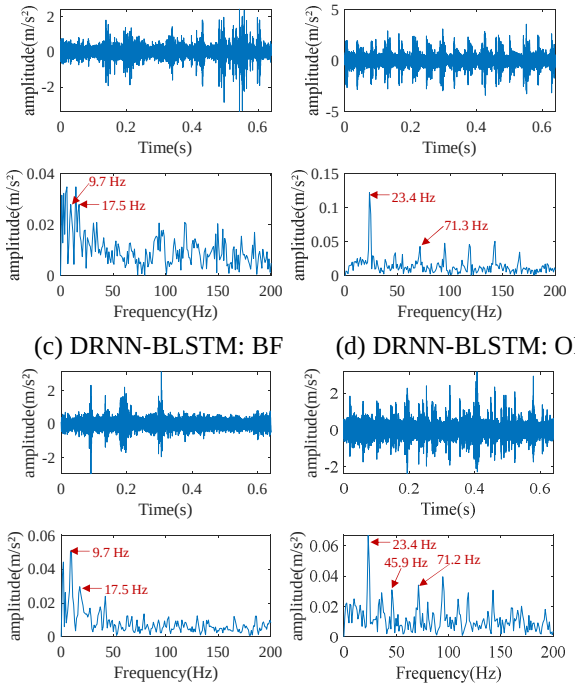
4.2. Comparative Experiments

To verify the advancement and superiority of the proposed model, we introduced two recently published state-of-the-art methods for comparison: the sparse decomposition-based compound fault feature separation method DASD-TIS (He, Li, Ding, & Zhang, 2022) and the deep learning-based single-channel blind source separation method DRNN-BLSTM (Issa & Al-Irhaym, 2021). DASD-TIS achieves blind separation of single-channel compound fault signals through dual-dictionary sparse decomposition and time-domain impulse separation, making it particularly suitable for separating overlapping faults under strong background noise. DRNN-BLSTM employs bidirectional LSTM networks to predict optimal ratio masks, enabling end-to-end training without requiring manual dictionary design.

The separation signals obtained by applying both comparative methods and the proposed method to the two decoupling tasks are shown in Figures 5 and Figure 6, where gear fault separation signals are displayed as Fourier spectra while bearing fault signals are shown as envelope spectra.



(a) DASD-TIS: BF (b) DASD-TIS: OR



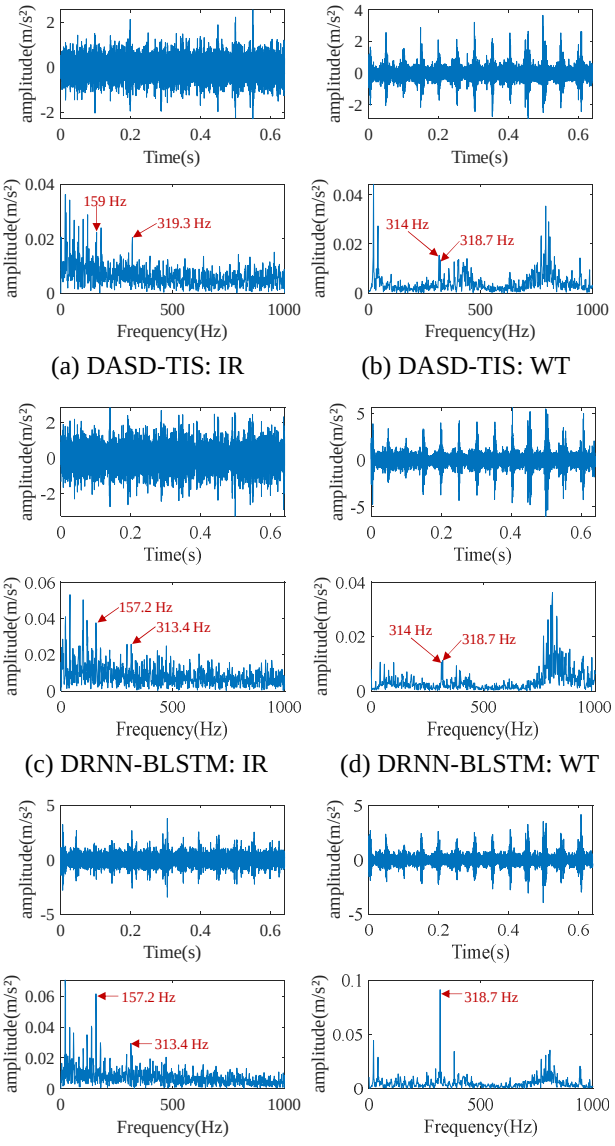
(c) DRNN-BLSTM: BF (d) DRNN-BLSTM: OR

(e) MVD-Net: BF (f) MVD-Net: OR

Figure 5 The time-domain waveform and envelope spectrum of the separated signal for Task A

In Task A, the envelope spectrum of DASD-TIS's BA separation signal showed no first-order FCF, only second-order FCF, while its OR separation signal envelope spectrum clearly displayed first and third-order FCFs. Although DRNN-BLSTM's BA separation signal envelope spectrum restored first-order FCF, it contained significant surrounding noise, while its OR separation signal envelope spectrum also showed relatively clear first and third-order FCFs. In comparison, MVD-Net's BA separation signal envelope spectrum clearly exhibited both first and second-order FCFs, and its OR separation signal envelope spectrum distinctly showed first, second and third-order FCFs. Clearly, in Task A, MVD-Net recovered more frequency details more accurately, producing results closer to the source signals.

In Task B, although DASD-TIS's IR separation signal envelope spectrum showed first and second-order FCFs, it contained substantial noise, while its WT separation signal spectrum only displayed relatively weak first-order FCF along with second-order FCF from IR, indicating incomplete separation of the WT signal. DRNN-BLSTM's separation results were similar to DASD-TIS, with its IR envelope spectrum containing significant noise components and its WT spectrum retaining IR information. In contrast, MVD-Net's IR envelope spectrum and WT spectrum both extracted pure FCFs with significantly higher signal-to-noise ratio than the comparative methods.



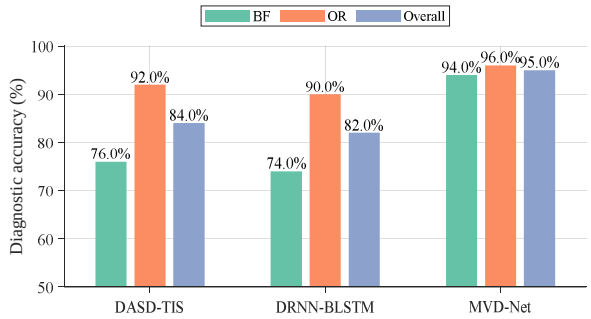
(a) DASD-TIS: IR (b) DASD-TIS: WT

(c) DRNN-BLSTM: IR (d) DRNN-BLSTM: WT

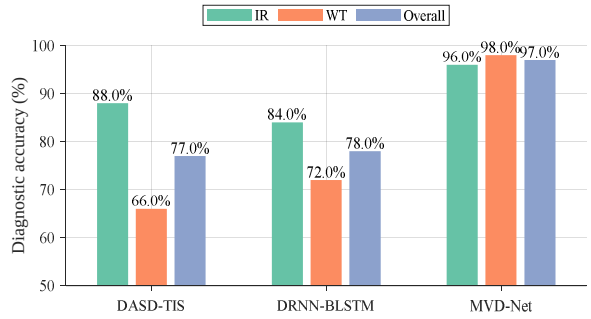
(e) MVD-Net: IR (f) MVD-Net: WT

Figure 6 The time-domain waveform and envelope spectrum (or spectrum) of the separated signal for Task B

The separation performance was evaluated based on fault diagnosis accuracy. A simple one-dimensional convolutional neural network-based diagnostic model was constructed. After training with the same single-fault training set, the diagnostic tests were performed on the separated signals obtained by the three methods. For each method, 50 separation experiments were conducted to obtain 50 sets of diagnostic samples. The diagnostic accuracy for each channel and the overall diagnostic accuracy are shown in Figure 7.



(a) Task A



(b) Task B

Figure 7 The diagnostic accuracy of the two tasks

Evidently, in both tasks, the separated signals obtained by the proposed MVD-Net method achieved higher diagnostic accuracy compared to the comparative methods. In Task A, while the diagnostic accuracy of OR separated signals by both DASD-TIS and DRNN-BLSTM was no less than 90%, the accuracy for BF separated signals was only 76% and 74% respectively. In contrast, MVD-Net achieved diagnostic accuracy above 90% for both channel-separated signals and the overall accuracy, demonstrating its advantage in separating weaker components in compound faults. In Task B, MVD-Net's diagnostic accuracy for both channel-separated signals and the overall accuracy exceeded 95%, significantly outperforming the comparative methods. This benefits from MVD-Net's capability to determine the order and number of separated signals, thereby enhancing the fault semantics of each channel. Additionally, the results also indicate that the proposed method is applicable to different components, showing good generalizability.

5. CONCLUSION

This study proposes a semi-supervised multi-source variational decoupling network for compound fault diagnosis of bogies under zero-shot conditions. The method requires only single-fault samples for training to achieve data-driven semi-blind source separation. The encoder extracts low-dimensional features from compound fault signals, and variational inference maps these features to multiple independent latent spaces corresponding to different fault sources, automatically determining the number of fault sources under the effect of ELBO. The designed independent decoders and source-selective activation mechanism ensure each separation channel focuses on one fault type, effectively alleviating the random semantic assignment problem caused by blind source separation ordering uncertainty. Experimental results show that compared with baseline methods, the proposed approach achieves better separation of compound fault signals in both axle boxes and gearboxes, while obtaining higher accuracy in fault diagnosis tasks.

REFERENCES

Cai, B., & Tang, G. (2024). Maximum spectral sparse entropy blind deconvolution for bearing fault diagnosis. *IEEE Sensors Journal*, vol. 24(5), pp. 6451-6468. doi: 10.1109/JSEN.2023.3348148

Dong, W., Feng, B., Liu, Y., & Zhang, S. (2025). Adaptive blind deconvolution decomposition and its application in composite fault diagnosis of rolling bearings. *Scientific Reports*, vol. 15(1), pp. 15169. doi: 10.1038/s41598-025-99913-w

He, D., Wu, J., Jin, Z., Huang, C., Wei, Z., & Yi, C. (2025). AGFCN: A bearing fault diagnosis method for high-speed train bogie under complex working conditions. *Reliability Engineering & System Safety*, vol. 258, pp. 110907. doi: 10.1016/j.ress.2025.110907

Huang, R., Liao, Y., Zhang, S., & Li, W. (2018). Deep decoupling convolutional neural network for intelligent compound fault diagnosis. *IEEE Access*, vol. 7, pp. doi: 1848-1858. 10.1109/ACCESS.2018.2886343

He, Z., Chu, P., Li, C., Zhang, K., Wei, H., & Hu, Y. (2023). Compound fault diagnosis for photovoltaic arrays based on multi-label learning considering multiple faults coupling. *Energy Conversion and Management*, vol. 279, pp. doi: 116742. 10.1016/j.enconman.2023.116742

He, G., Li, J., Ding, K., & Zhang, Z. (2022). Feature extraction of gear and bearing compound faults based on vibration signal sparse decomposition. *Applied acoustics*, vol. 189, pp. 108604. doi: 10.1016/j.apacoust.2021.108604

Issa, R. J., & Al-Irhaym, Y. F. (2021). Audio source separation using supervised deep neural network. *In Journal of Physics: Conference Series* (Dec 9-10), Bagdad, Iraq. doi: 10.1088/1742-6596/1879/2/022077

Jiang, Y., Zhu, H., & Li, Z. (2016). A new compound faults detection method for rolling bearings based on empirical wavelet transform and chaotic oscillator. *Chaos, Solitons & Fractals*, vol. 89, pp. 8-19. doi: 10.1016/j.chaos.2015.09.007

Jin, Y., Qin, C., Huang, Y., & Liu, C. (2021). Actual bearing compound fault diagnosis based on active learning and decoupling attentional residual network. *Measurement*, vol. 173, pp. 108500. doi: 10.1016/j.measurement.2020.108500

Lyu, X., Hu, Z., Zhou, H., & Wang, Q. (2019). Application of improved MCKD method based on QGA in planetary

gear compound fault diagnosis. *Measurement*, vol. 139, pp. 236-248. doi: 10.1016/j.measurement.2019.02.071

Peng, D., Zhu, X., Teng, W., & Liu, Y. (2023). Use of generalized Gaussian cyclostationarity for blind deconvolution and its application to bearing fault diagnosis under non-Gaussian conditions. *Mechanical Systems and Signal Processing*, vol. 196, pp. 110351. doi: 10.1016/j.ymssp.2023.110351

Wan, S., Zhang, X., & Dou, L. (2018). Compound fault diagnosis of bearings using improved fast spectral kurtosis with VMD. *Journal of Mechanical Science and Technology*, vol. 32(11), pp. 5189-5199. doi: 10.1007/s12206-018-1017-8

Xiao, C., & Yu, J. (2023). Adaptive swarm decomposition algorithm for compound fault diagnosis of rolling bearings. *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-14. doi: 10.1109/TIM.2022.3231324

Zhang, W. T., Ji, X. F., Huang, J., & Lou, S. T. (2021). Compound fault diagnosis of aero-engine rolling element bearing based on CCA blind extraction. *IEEE Access*, vol. 9, pp. 159873-159881. doi: 10.1109/ACCESS.2021.3130637

Zhang, Z., Li, S., Xin, Y., & Ma, H. (2020). A novel compound fault diagnosis method using intrinsic component filtering. *Measurement Science and Technology*, vol. 31(5), pp. 055103. doi: 10.1088/1361-6501/ab62c5

Zhong, J, Yang, Z and Wong, S. F. (2010). Machine condition monitoring and fault diagnosis based on support vector machine. *2010 IEEE International Conference on Industrial Engineering and Engineering Management* (2228-2233), Oct 29-31, Macao, China. doi: 10.1109/IEEM.2010.5674594.