Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 2, No. 2, 2026

DFPSE: Diagnostic-Feature-Preserving Signal Enhancement for Robust Bearing Fault Diagnosis under Compound Noise

Jiajun Liu [a], Baishun Su [b, *]

School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China
[a] 212409010018@home.hpu.edu.cn, [b] subaishun@hpu.edu.cn

Abstract: Rolling bearing fault diagnosis using deep learning (DL) achieves high accuracy under clean laboratory conditions, yet performance collapses under the compound noise typical of industrial environments — Gaussian white noise, harmonic interference, transient impulses, and sensor drift. Existing signal denoising methods optimize for waveform reconstruction fidelity, but we demonstrate a critical and counterintuitive finding: higher signal-to-noise ratio (SNR) does not imply better diagnostic accuracy, and reconstruction-only enhancement can actively degrade classification performance below the unenhanced noisy baseline. To address this disconnect, we propose Diagnostic-Feature-Preserving Signal Enhancement (DFPSE), a framework in which a frozen, pretrained diagnostic model (teacher) guides a lightweight Residual Temporal Convolutional Network (Residual TCN) enhancer via three complementary losses: waveform reconstruction, diagnostic classification, and feature-level alignment. Trained under a single Gaussian drift noise condition (SNR in [-10, 8] dB), the enhancer is evaluated across six unseen noise conditions spanning five Gaussian drift levels and a compound interference scenario. Experiments on the Huazhong University of Science and Technology (HUST) and Paderborn University (PU) bearing datasets with three architecturally diverse teachers — ResNet1D (CNN), TCN (dilated convolution), and Vanilla Transformer (self-attention) — demonstrate that DFPSE consistently recovers diagnostic accuracy with gains of 22–29 percentage points on HUST (average across conditions) and 9–22 points on PU. Critically, on the HUST dataset at -2 dB input SNR, the reconstruction-only variant achieves 5.59 dB output SNR but only 65.4% accuracy (below the 86.7% noisy baseline), whereas DFPSE achieves 2.72 dB SNR with 98.4% accuracy — a 33-point gap that reveals a fundamental trade-off between waveform fidelity and diagnostic discriminability. This finding challenges the conventional assumption that better denoising necessarily yields better diagnosis and argues for task-aware evaluation in signal enhancement research.

Keywords: Bearing fault diagnosis, signal enhancement, knowledge distillation, compound noise, deep learning.

1. Introduction

1.1. Background and Motivation

Bearing failures account for a substantial fraction of rotating machinery breakdowns in industrial systems [1]. In recent years, DL methods have been widely applied to vibration-based bearing fault diagnosis, with convolutional, recurrent, and transformer-based architectures reporting classification accuracies above 99% on benchmark datasets under controlled laboratory conditions [1, 2]. Convolutional neural networks, recurrent architectures, autoencoders, and transformer-based models have demonstrated classification accuracies exceeding 99% on benchmark datasets under controlled laboratory conditions [1, 2]. These successes have driven intensive research toward deploying DL-based diagnostic systems in real-world industrial settings.

However, a critical gap persists between laboratory benchmarks and industrial reality. In practical environments, vibration signals are corrupted by compound noise that includes Gaussian white noise, harmonic interference from adjacent machinery, random transient impulses, and sensor drift — a far more complex degradation than the simple additive Gaussian noise assumed in most studies. Comprehensive surveys consistently identify noise robustness as the primary barrier preventing the industrial deployment of DL-based fault diagnosis systems [3, 2]. When exposed to compound noise, the diagnostic accuracy of even state-of-the-art models can drop by tens of percentage points, rendering them unfit for deployment in noise-prone industrial settings.

1.2. Existing Approaches and Their Limitations

Three broad families of approaches have been developed to address noise in bearing fault diagnosis.

Classical signal processing methods employ spectral analysis, wavelet transforms, and adaptive mode decomposition to separate fault-related components from noise [4]. While interpretable and well-established, these methods rely on hand-crafted features and prior knowledge of fault characteristic frequencies, limiting their adaptability to complex machinery and varying operating conditions.

Transfer learning and domain adaptation approaches aim to align feature distributions between source (clean or known-condition) and target (noisy) domains [5, 6]. These methods are effective when the domain shift is primarily distributional — for instance, across different operating loads or speeds — but they do not directly address the problem of raw signal corruption by noise. When the target signal is severely degraded, feature alignment alone cannot recover lost diagnostic information.

Joint denoising and diagnosis frameworks train neural networks to simultaneously suppress noise and perform fault classification [7, 8]. These methods represent a significant advance over pipeline approaches that treat denoising and diagnosis as independent stages. The attention-guided joint

learning CNN proposed in [7] demonstrates that coupling denoising with diagnosis objectives improves robustness, while [8] shows that adaptive 1D-CNN structures can unify both tasks within a single architecture.

Despite their contributions, all three families share a common, largely unexamined assumption: that improving waveform-level signal quality — measured through SNR, root-mean-square error (RMSE), or similar reconstruction metrics — necessarily leads to better diagnostic performance. This assumption has guided the design of denoising objectives toward waveform fidelity, implicitly treating the denoising problem as one of signal reconstruction.

1.3. The Core Challenge: Reconstruction Quality vs. Diagnostic Discriminability

Recent work has begun to challenge this assumption. The Denoising Fault-Aware Wavelet Network (DFAWNet) [9] incorporates signal processing priors into the network architecture to preserve fault-related features during denoising, suggesting that what is preserved matters more than how much is removed. Physics-informed time-frequency fusion networks [10] similarly demonstrate that guiding denoising with domain knowledge improves diagnostic outcomes. Meanwhile, dual-domain adversarial methods [11] operate in both time and frequency domains to target noise more precisely. These advances point toward a critical insight: the objective of signal enhancement for diagnosis should not be waveform reconstruction but diagnostic feature preservation. Yet, no existing framework systematically decouples these two objectives — nor does it provide a mechanism for diagnostic models to directly supervise the signal enhancement process.

This limitation motivates a fundamentally different approach. If the ultimate goal of signal enhancement is to improve diagnostic accuracy rather than to faithfully reconstruct clean waveforms, then the enhancement process should be guided directly by diagnostic objectives. In other words, we should let the diagnostic model itself define what constitutes a "good" enhanced signal.

1.4. DFPSE: Proposed Framework

In this paper, we propose DFPSE, a framework that operationalizes this insight through a teacher-student knowledge distillation (KD) paradigm. The core idea is straightforward: a frozen, pretrained diagnostic model (teacher) that achieves high accuracy on clean signals is used to guide the training of a lightweight Residual TCN enhancer. The enhancer learns to transform noisy signals into enhanced versions that the teacher can classify correctly — even if those enhanced signals do not closely resemble the original clean waveforms in the waveform sense.

DFPSE employs three complementary loss functions to realize this objective:

Reconstruction loss ($\mathcal{L}_{\text{rec}}$): a smooth L1 loss between the enhanced and clean signals, providing a waveform-level anchor to prevent the enhancer from producing physically meaningless outputs.

Classification loss ($\mathcal{L}_{\text{cls}}$): the cross-entropy loss of the frozen teacher on the enhanced signal, directly optimizing for diagnostic accuracy.

Feature-alignment loss ($\mathcal{L}_{\text{feat}}$): a normalized L2 distance between the teacher's intermediate feature representations of the clean and enhanced signals, ensuring that the enhanced signal activates the same diagnostic pathways as the clean signal.

Critically, the teacher is never updated during enhancer training — its parameters remain frozen, serving as a fixed diagnostic oracle. This design ensures that DFPSE is agnostic to the teacher architecture and can be applied with any pretrained diagnostic model.

1.5. Contributions

The main contributions of this work are as follows:

A diagnostic-guided signal enhancement framework. DFPSE decouples the objectives of signal enhancement and fault diagnosis while establishing a principled mechanism — frozen-teacher KD — through which diagnostic requirements directly supervise the enhancement process. Unlike prior joint denoising-diagnosis approaches, DFPSE does not require retraining the diagnostic model.

Empirical discovery of the reconstruction-diagnosis trade-off. Through systematic ablation studies, we demonstrate that the reconstruction-only variant of DFPSE achieves higher SNR but significantly lower diagnostic accuracy than the full DFPSE framework. This finding reveals a fundamental tension between waveform fidelity and diagnostic discriminability — a result that challenges the conventional assumption underlying most signal denoising research for fault diagnosis.

Train-once, evaluate-many generalization. The enhancer is trained under a single noise condition (time-varying Gaussian drift, SNR in [-10, 8] dB) and evaluated across six distinct noise conditions spanning Gaussian drift at five SNR levels and a compound interference scenario. DFPSE consistently recovers diagnostic accuracy without per-condition retraining or noise-type knowledge at test time.

Cross-architecture validation. We validate DFPSE with three architecturally diverse teachers — ResNet1D (CNN family), TCN (dilated convolution family), and Vanilla Transformer (attention family) — on two public bearing datasets. The consistent performance gains across all architectures confirm the framework's generality.

1.6. Paper Organization

The remainder of this paper is organized as follows. Section 2 reviews related work on signal enhancement, KD, and diagnostic feature preservation. Section 3 presents the DFPSE methodology, including the enhancer architecture, loss formulation, and training protocol. Section 4 describes the experimental setup, covering datasets, teacher models, noise configurations, and evaluation metrics. Section 5 presents and discusses the experimental results, including ablation studies, SNR-accuracy analysis, robustness evaluation, and cross-architecture validation. Section 6 concludes the paper and outlines directions for future work.

2. Related Work

This section reviews three lines of research most closely related to DFPSE: signal denoising methods for fault diagnosis, KD paradigms, and joint denoising-diagnosis frameworks. For each, we identify the advances made and the limitations that motivate our approach.

2.1. Signal Denoising for Fault Diagnosis

Vibration signals from industrial bearings are inevitably contaminated by noise from electromagnetic interference, adjacent machinery, and sensor artifacts. Restoring diagnostic information from noisy signals has been a central research

challenge for decades.

Early approaches relied on classical signal processing: wavelet thresholding, singular value decomposition (SVD), empirical mode decomposition (EMD), and spectral subtraction [4]. These methods exploit known properties of fault signals — such as characteristic frequencies and impulsive signatures — to separate fault-related components from noise. While interpretable and computationally efficient, they require careful hand-tuning of parameters (e.g., decomposition levels, threshold values) and prior knowledge of bearing geometry and operating speed. Their performance degrades when fault signatures overlap with noise in the frequency domain, as is common under compound interference.

The advent of DL brought data-driven alternatives. Denoising autoencoders (DAEs), convolutional autoencoders, and generative adversarial networks (GANs) learn noise-suppression mappings directly from paired clean-noisy signal examples. Wang et al. [7] proposed an attention-guided joint learning CNN that couples denoising and diagnosis within a single architecture, using attention to focus the network on fault-related signal components. Han et al. [11] introduced a dual-domain adversarial method operating in both time and frequency domains, demonstrating that multi-domain information improves denoising fidelity. Kim and Kim [10] incorporated physics-informed priors into time-frequency fusion networks, showing that domain knowledge constrains the learning process toward physically meaningful solutions.

A parallel development is self-supervised and blind denoising. Unlike conventional supervised methods that require paired clean-noisy data, self-supervised approaches such as Noise2Noise [12] learn to denoise without explicit clean references, making them attractive for real-world scenarios where clean baselines are unavailable. Van den Ende et al. [12] demonstrated a self-supervised blind denoising framework for distributed acoustic sensing that achieves strong waveform coherence without ground-truth clean signals.

Despite their technical diversity, all the above approaches share a common design principle: they optimize for waveform-level fidelity, measured by metrics such as SNR, RMSE, or perceptual quality. None incorporates the downstream diagnostic task as a supervision signal for the denoising process. This is the central limitation that DFPSE addresses.

2.2. Knowledge Distillation

Knowledge Distillation , originally proposed by Hinton et al. for model compression [13], transfers the predictive capability of a large teacher model to a smaller student model by aligning their output distributions. The paradigm has since been extended along multiple dimensions, including feature-level distillation that matches intermediate representations [14] and multi-teacher distillation that aggregates knowledge from diverse architectures.

In the fault diagnosis domain, KD has been adopted primarily for two purposes: model compression for edge deployment and cross-domain knowledge transfer. Lu et al. [15] designed a lightweight KD-based transfer learning framework where a deep variable-scale residual teacher trains a compact student for bearing diagnosis across devices, achieving high accuracy with significantly reduced parameters. Zhou et al. [16] combined federated transfer learning with multi-teacher KD, where knowledge from multiple source domains is dynamically weighted and distilled to a target model — demonstrating that KD can aggregate heterogeneous diagnostic knowledge under privacy constraints.

A key observation from this body of work is that the teacher in existing KD pipelines is always a classifier whose knowledge is transferred to another classifier. The teacher's role is exclusively to produce soft labels or feature representations for the student to mimic. No existing work, to our knowledge, has explored the inverse paradigm: using a frozen diagnostic teacher to guide a signal enhancer — a model whose output is not a class prediction but an enhanced signal intended for subsequent diagnosis. DFPSE occupies this unexplored intersection.

2.3. Joint Denoising-Diagnosis Frameworks

A natural strategy for noise-robust diagnosis is to unify denoising and diagnosis within a single network, eliminating the information bottleneck of pipeline approaches where denoising artifacts can propagate to the classifier. Several recent works have pursued this direction.

Hong and Kim [8] proposed an adaptive 1D-CNN structure in which denoising and diagnostic layers are jointly optimized, showing that end-to-end training improves both tasks compared to sequential processing. The attention-guided joint learning CNN of Wang et al. [7] further refines this idea by applying attention mechanisms that direct the denoising subnet toward fault-relevant temporal regions.

Beyond pure data-driven unification, signal-processing-informed architectures incorporate domain knowledge into network design. Shang et al. [9] developed the Denoising Fault-Aware Wavelet Network (DFAWNet), which embeds wavelet decomposition into the neural architecture to preserve fault-related features during denoising. Their results suggest that what is preserved during denoising is often more important than how much noise is removed. This finding aligns with the motivating observation of DFPSE: waveform reconstruction quality and diagnostic discriminability are distinct objectives that can conflict.

However, even joint frameworks treat the denoising module as a generic preprocessing stage that shares features with the classifier. They do not provide an explicit mechanism for a diagnostic model to define what constitutes adequate signal enhancement. The denoiser learns to produce signals that the classifier can process, but it lacks direct diagnostic supervision — it does not receive gradient signals indicating which temporal patterns are discriminatively important.

2.4. Positioning of DFPSE

DFPSE differs from the surveyed methods in three fundamental respects. First, it inverts the conventional KD paradigm: the teacher is not compressed into a smaller classifier but used as a frozen diagnostic oracle that supervises the training of a signal enhancer. Second, through the feature-alignment loss, DFPSE explicitly transfers diagnostic knowledge at the level of intermediate representations, ensuring that the enhanced signal activates the same diagnostic pathways as the clean signal — a constraint absent in joint denoising-diagnosis networks. Third, the enhancer is trained under a single noise condition and evaluated across six unseen conditions, demonstrating generalization that goes beyond the per-condition retraining typical of prior denoising work.

Taken together, these distinctions place DFPSE in a

different category from prior denoising and KD methods: the enhancement objective is defined by a frozen diagnostic model rather than by waveform similarity, and the framework is agnostic to the teacher architecture, making it applicable to any diagnostic model pretrained on clean data.

3. Methodology

This section presents the DFPSE framework. We first formalize the problem setting, then describe the enhancer architecture, the three complementary loss functions, and the training and evaluation protocol.

3.1. Problem Formulation

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ denote a dataset of clean vibration signals $\mathbf{x}_i \in \mathbb{R}^L$ and their corresponding fault labels $y_i \in \{1, \dots, K\}$, where L is the signal length and K is the number of fault categories. We assume access to a pretrained diagnostic model $f_\theta: \mathbb{R}^L \to \mathbb{R}^K$ (the teacher) that achieves high classification accuracy on $\mathcal{D}$ under clean conditions. The teacher parameters θ are frozen and never updated during enhancer training.

In real industrial environments, the measured signal is corrupted by compound noise:

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i + \boldsymbol{\eta}_i,$$

Where $\boldsymbol{\eta}_i$ represents a combination of Gaussian white noise, harmonic interference, random transient impulses, and sensor drift. When $\tilde{\mathbf{x}}_i$ is directly fed into f_θ, classification accuracy degrades sharply because the noise masks the diagnostic features that the teacher learned from clean signals.

Our goal is to train a signal enhancer $g_\phi: \mathbb{R}^L \to \mathbb{R}^L$ parameterized by ϕ, such that

$$f_\theta\left(g_\phi(\tilde{\mathbf{x}}_i)\right) \approx f_\theta(\mathbf{x}_i),$$

i.e., the enhanced signal $\hat{\mathbf{x}}_i = g_\phi(\tilde{\mathbf{x}}_i)$ preserves the diagnostic features needed for accurate classification. Critically, we do not require $\hat{\mathbf{x}}_i \approx \mathbf{x}_i$ in the waveform sense — the enhancer is free to produce signals that deviate from the clean waveform as long as they are diagnostically discriminative.

3.2. Enhancer Architecture

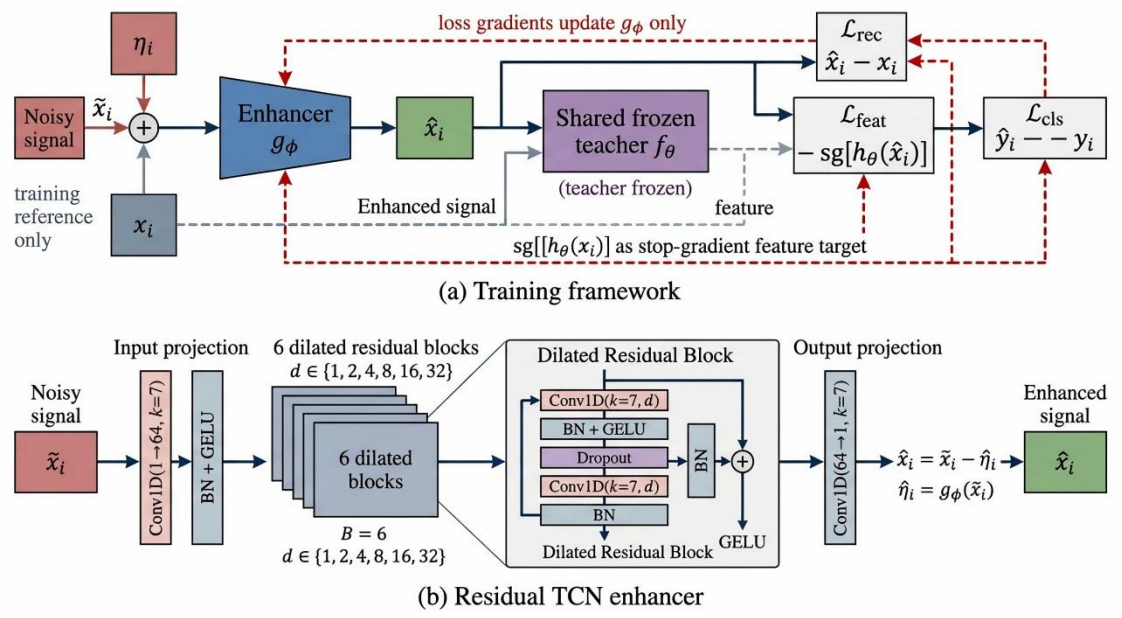

Figure 1. DFPSE framework overview. The Residual TCN enhancer maps noisy signals to enhanced signals under supervision from a frozen diagnostic teacher using reconstruction, classification, and feature-alignment losses. $\mathcal{L}_{rec}\mathcal{L}_{cls}\mathcal{L}_{feat}$

The enhancer g_ϕ adopts a Residual TCN architecture, chosen for its efficiency and ability to capture multi-scale temporal dependencies. As illustrated in Fig. 1, the architecture consists of three stages:

Input projection. The noisy 1D signal $\tilde{\mathbf{x}} \in \mathbb{R}^L$ is first projected to a C-dimensional feature space through a 1D convolution with kernel size k_p, followed by batch normalization and GELU activation:

$$\mathbf{z}_0 = \mathrm{GELU}\left(\mathrm{BN}\left(\mathrm{Conv1D}(\tilde{\mathbf{x}}; C, k_p)\right)\right) \in \mathbb{R}^{C \times L}$$

Dilated residual blocks. The projected features pass through a stack of B dilated residual blocks, each operating at the same channel dimension C. Block b uses a dilation factor d_b that grows exponentially across blocks ($d \in \{1,2,4,8,16,32\}$), enabling the receptive field to expand without increasing parameter count. Each block consists of two 1D convolutional layers with kernel size k_r, batch normalization, GELU activations, and dropout regularization, wrapped in a residual connection:

$$\mathbf{z}_b = \mathbf{z}_{b-1} + \mathcal{F}_b(\mathbf{z}_{b-1}; d_b)$$

Where $\mathcal{F}_b$ denotes the stacked convolution-batchnorm-activation operations within block b. The residual structure stabilizes training and preserves fine-grained temporal information through the stack.

Output projection and residual learning. The final features are mapped back to the signal space via a 1×1 convolution, producing a noise estimate $\hat{\boldsymbol{\eta}}$. The enhanced signal is obtained through residual subtraction:

$$\hat{\mathbf{x}} = \tilde{\mathbf{x}} - \hat{\boldsymbol{\eta}}$$

Residual learning is employed because the noise component is typically easier to estimate than the clean signal, and this formulation naturally preserves the original signal structure.

In our implementation, we set $C = 64$, $k_p = k_r = 7$, $B = 6$, and dropout rate 0.0 (training regularization is achieved

through the noise diversity in the data). The resulting model contains approximately 85K parameters, making it lightweight and suitable for deployment even on resource-constrained edge devices.

3.3. Loss Functions

DFPSE trains the enhancer through three complementary loss functions that jointly encode the objectives of waveform anchoring, diagnostic accuracy, and feature-level alignment.

3.3.1. Reconstruction Loss

The reconstruction loss provides a waveform-level anchor that prevents the enhancer from producing physically meaningless outputs. We employ the Smooth L1 (Huber) loss between the enhanced and clean signals:

$$\mathcal{L}_{\text{rec}} = \frac{1}{N}\sum_{i=1}^{N} \text{SmoothL1}\left(\hat{\mathbf{x}}_i, \mathbf{x}_i\right)$$

Where $\text{SmoothL1}(a, b) = \frac{1}{L}\sum_j \ell\left(a_j - b_j\right)$ with $\ell(z) = 0.5z^2$ for $|z| < 1$ and $|z| - 0.5$ otherwise. Smooth L1 is preferred over MSE because it is less sensitive to large residuals caused by transient noise impulses.

3.3.2. Classification Loss

The classification loss directly optimizes the enhancer for diagnostic accuracy. The enhanced signal is passed through the frozen teacher, and the cross-entropy between the teacher's prediction and the ground-truth label is minimized:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp\left(f_\theta^{(y_i)}(\hat{\mathbf{x}}_i)\right)}{\sum_{k=1}^{K} \exp\left(f_\theta^{(k)}(\hat{\mathbf{x}}_i)\right)}$$

Where $f_\theta^{(k)}(\cdot)$ denotes the logit for class k. Because θ is frozen, gradients from $\mathcal{L}_{\text{cls}}$ flow exclusively to the enhancer parameters ϕ, forcing the enhancer to produce signals that the teacher already knows how to classify.

3.3.3. Feature-Alignment Loss

The feature-alignment loss operates at a deeper level than the classification loss. Beyond producing correct class predictions, we want the enhanced signal to activate the same intermediate representations in the teacher as the clean signal. Let $\mathbf{h}_\theta(\mathbf{x}) \in \mathbb{R}^D$ denote the penultimate-layer feature vector of the teacher. The feature-alignment loss is the normalized L2 distance:

$$\mathcal{L}_{\text{feat}} = \frac{1}{N}\sum_{i=1}^{N} \left\| \frac{\mathbf{h}_\theta(\hat{\mathbf{x}}_i)}{\|\mathbf{h}_\theta(\hat{\mathbf{x}}_i)\|_2} - \frac{\mathbf{h}_\theta(\mathbf{x}_i)}{\|\mathbf{h}_\theta(\mathbf{x}_i)\|_2} \right\|_2^2$$

Normalization to unit length removes scale differences and focuses the loss on angular alignment in the feature space. The clean-signal features $\mathbf{h}_\theta(\mathbf{x}_i)$ are detached from the computation graph, serving as fixed targets.

3.3.4. Total Loss and Ablation Variants

The total training loss is a weighted sum:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} + \lambda_{\text{feat}}\mathcal{L}_{\text{feat}}$$

Where λ_{rec}, λ_{cls}, and λ_{feat} are scalar weights. In our experiments, we set $\lambda_{\text{rec}} = 1.0$, $\lambda_{\text{cls}} = 0.2$, and $\lambda_{\text{feat}} = 0.5$.

To isolate the contribution of each loss component, we define three ablation variants:

Rec-Only ($\lambda_{\text{cls}} = \lambda_{\text{feat}} = 0$): optimizes solely for waveform reconstruction fidelity.

Rec+Cls ($\lambda_{\text{feat}} = 0$): adds classification supervision to reconstruction.

DFPSE (Full): employs all three losses as defined above.

Comparing Rec-Only against DFPSE allows us to quantify the impact of diagnostic guidance on signal enhancement — and to test the hypothesis that better waveform reconstruction does not necessarily imply better diagnosis.

3.4. Training and Evaluation Protocol

3.4.1. Train-Once, Evaluate-Many Paradigm

A key design principle of DFPSE is that the enhancer should generalize across noise conditions without per-condition retraining. We adopt a train-once, evaluate-many protocol:

Training noise condition: Time-varying Gaussian drift with SNR uniformly sampled from [-10, 8] dB. The SNR drifts slowly over the signal duration, simulating gradual changes in environmental noise intensity.

Test noise conditions (6 total): Five levels of Gaussian drift at fixed SNR values (-10, -8, -6, -4, -2 dB) plus one compound interference condition that combines Gaussian white noise, harmonic tones at random frequencies, random transient impulses, and baseline drift.

At test time, the enhancer receives noisy signals from conditions it was never trained on and must produce enhanced signals that the frozen teacher can classify. No noise-type information is provided to the enhancer.

3.4.2. Teacher Models

To demonstrate the generality of DFPSE, we evaluate with three architecturally diverse teacher models spanning the major DL paradigms for 1D signal classification:

ResNet1D (241K parameters): A 1D residual convolutional network with base channel width 32 and dropout 0.1. Represents the CNN family.

TCN (173K parameters): A temporal convolutional network with channel progression (32, 64, 128), kernel size 5, and dropout 0.2. Represents the dilated convolution family.

Vanilla Transformer (403K parameters): A standard Transformer encoder with $d_{\text{model}} = 128$, 4 heads, 4 layers, and feedforward dimension 256. Input signals are downsampled by factor 8 before positional encoding. Represents the self-attention family.

All teachers are pretrained on clean signals from the respective datasets and achieve the clean-condition accuracies reported in Section 4. During enhancer training, teacher parameters are frozen (requires_grad=False) and used exclusively for loss computation.

3.4.3. Optimization

The enhancer is trained using the AdamW optimizer with a learning rate of 10^{-3}, weight decay 10^{-4}, and batch size 64. Training proceeds for 50 epochs. All experiments use a fixed random seed (18) for reproducibility.

4. Experimental Setup

This section describes the datasets, teacher models, noise configurations, evaluation metrics, and implementation details used in all experiments.

4.1. Datasets

We evaluate DFPSE on two publicly available rolling bearing fault datasets that differ in sensor type, operating conditions, and fault characteristics, providing a rigorous test of the framework's generality.

4.1.1. PU Dataset

The PU bearing dataset contains vibration signals collected from a modular test rig with healthy and artificially damaged bearings. We use 10 fault categories: one healthy class (K001) and nine damage types spanning inner race faults (KI14, KI16, KI17), outer race faults (KA04, KA16, KA30), and ball faults (KB23, KB24, KB27) across different damage severities and manufacturing methods (electric discharge machining, drilling, and manual engraving). Signals are segmented into fixed-length windows of $L = 8192$ samples. The dataset is randomly partitioned into 9,863 training, 3,283 validation, and 3,294 test windows with no overlap between splits.

4.1.2. HUST Dataset

The HUST bearing dataset is collected from a rotor-bearing testbed under varying speeds (600–3000 rpm) and includes both healthy and faulty conditions. We use 9 categories: healthy (H) and eight fault types across inner race, outer race, ball, and cage faults at different severity levels (denoted 0.5X and 1X damage width). Window length is set to $L = 4096$ samples. The dataset is partitioned into 7,425 training, 2,376 validation, and 2,376 test windows.

HUST serves as the primary evaluation dataset because its clean-signal diagnostic accuracy leaves substantial room for improvement — the best teacher achieves only 83.8% on clean signals, compared to over 99% on PU. This makes HUST the more challenging and diagnostically meaningful benchmark for evaluating signal enhancement.

4.2. Teacher Models

We employ three pretrained teacher models spanning the major architectural paradigms for 1D time-series classification. All teachers are trained exclusively on clean signals from the respective datasets and are frozen during enhancer training. Their clean-condition test accuracy on both datasets is reported in Table 1.

Table 1. Teacher architectures and clean-signal accuracy

Teacher	Params	PU Acc. (%)	HUST Acc. (%)
ResNet1D (CNN)	241K	99.9	83.8
TCN (Dilated Conv)	173K	99.6	70.8
Vanilla Transformer	403K	84.5	71.8

ResNet1D is a 1D residual convolutional network with base channel width 32 and dropout rate 0.1. It employs residual blocks with skip connections and batch normalization, representing the CNN paradigm that has been widely adopted for vibration-based fault diagnosis.

TCN uses stacked 1D convolutions with exponentially increasing dilation factors, implemented as three stages with channel dimensions (32, 64, 128), kernel size 5, and dropout 0.2. The dilated architecture provides an exponentially growing receptive field while maintaining parameter efficiency.

Vanilla Transformer follows the standard Transformer encoder design with $d_{model} = 128$, 4 attention heads, 4 encoder layers, and feedforward dimension 256. Raw signals are downsampled by a factor of 8 through a strided convolution before positional encoding, reducing the sequence length to a computationally tractable size. Dropout of 0.2 is applied in attention and feedforward sub-layers.

4.3. Noise Configurations

4.3.1. Training Noise

The enhancer is trained exclusively under time-varying Gaussian drift noise. In this regime, the noise power varies smoothly over the signal duration, simulating gradual changes in environmental interference. Specifically, the instantaneous SNR at each time step drifts linearly within the range [-10, 8] dB, controlled by a drift speed parameter of 0.5. This configuration exposes the enhancer to a continuum of SNR levels during training without requiring discrete per-level data generation.

4.3.2. Test Noise Conditions

At test time, we evaluate the enhancer on six distinct noise conditions that were never seen during training:

Table 2. Test noise conditions

Condition	Noise setting
1	Gaussian drift at -10 dB (SNR ≈ -10 dB)
2	Gaussian drift at -8 dB (SNR ≈ -8 dB)
3	Gaussian drift at -6 dB (SNR ≈ -6 dB)
4	Gaussian drift at -4 dB (SNR ≈ -4 dB)
5	Gaussian drift at -2 dB (SNR ≈ -2 dB)
6	Compound severe (SNR ≈ -9 dB)

Conditions 1–5 systematically vary the noise intensity of the same distribution family used in training, testing the enhancer's ability to interpolate and extrapolate across SNR levels. Condition 6 introduces a qualitatively different noise composition, a composite interference consisting of Gaussian white noise, harmonic tones at random frequencies, random transient impulses, and baseline drift testing generalization to an unseen noise type.

4.4. Evaluation Metrics

We evaluate diagnostic performance using four complementary metrics:

Accuracy (Acc): overall classification accuracy on the test set, reported as a percentage.

Macro-averaged F1-score (mF1): the unweighted mean of per-class F1-scores, providing a balanced assessment across all fault categories.

Macro-averaged Recall (mRecall): the unweighted mean of per-class recall, measuring the framework's ability to detect each fault type.

Per-class Recall: recall for individual fault categories, used to diagnose category-specific weaknesses.

Additionally, we report SNR improvement and RMSE to quantify waveform-level signal quality of the enhanced output relative to the clean reference. SNR is computed as $\mathrm{SNR} = 10\log_{10}(P_{clean}/P_{error})$, where P_{error} is the power of the difference between the enhanced and clean signals.

4.5. Implementation Details

Enhancer configuration. The Residual TCN enhancer uses $C = 64$ hidden channels, kernel size $k_p = k_r = 7$, $B = 6$ dilated residual blocks with dilation factors $(1, 2, 4, 8, 16, 32)$, and dropout 0.0 (regularization is achieved through the diverse noise distribution during training). The model contains approximately 85K parameters.

Training protocol. The enhancer is trained for 100 epochs using the AdamW optimizer with learning rate 10^{-3}, weight decay 10^{-4}, and batch size 64. A cosine annealing schedule

reduces the learning rate to 10^{-6} over the course of training. Early stopping is applied with a patience of 15 epochs monitored on the validation classification loss. Loss weights are set to $\lambda_{rec} = 1.0$, $\lambda_{cls} = 0.1$, and $\lambda_{feat} = 0.1$.

Reproducibility. All experiments use a fixed random seed (18) for data loading and network initialization to ensure reproducibility.

Hardware and software. Experiments are conducted on a single NVIDIA GeForce RTX 4060 GPU (8GB VRAM) using PyTorch 2.7.0 with CUDA 12.6. The complete codebase, including model definitions, training scripts, and evaluation pipelines, is implemented in Python and will be publicly released upon acceptance.

5. Results and Discussion

This section presents a systematic evaluation of DFPSE. We first conduct ablation studies to isolate the contribution of each loss component and to test the central hypothesis that waveform reconstruction quality does not imply diagnostic discriminability. We then analyze robustness across noise intensities, validate cross-architecture generality on two datasets, examine per-class recovery through confusion matrices, and provide qualitative evidence through waveform and envelope spectrum visualization.

5.1. Ablation Study

To understand the role of each loss term, we compare three configurations on the HUST dataset with the ResNet1D teacher: (i) Noisy Baseline — the noisy signal is directly classified by the frozen teacher without any enhancement; (ii) Rec-Only — the enhancer is trained with only the reconstruction loss $\mathcal{L}_{rec}$; and (iii) DFPSE — the full framework with all three losses.

Table 3 reports accuracy, macro-F1, and SNR for all six test noise conditions. Figure 2 visualizes these results as grouped bar charts.

Table 3. Ablation results on HUST with ResNet1D

Noise	Metric	Noisy	Rec-Only	DFPSE	Clean
-10 dB	Acc	39.2	51.3	75.1	83.8
	mF1	0.358	0.511	0.752	0.833
	SNR	-10.0	0.74	-1.10	∞
-8 dB	Acc	52.3	56.1	87.3	83.8
	mF1	0.498	0.562	0.871	0.833
	SNR	-8.0	1.97	-0.01	∞
-6 dB	Acc	65.0	61.4	95.2	83.8
	mF1	0.634	0.614	0.952	0.833
	SNR	-6.0	3.17	0.96	∞
-4 dB	Acc	77.4	63.9	97.4	83.8
	mF1	0.767	0.637	0.974	0.833
	SNR	-4.0	4.38	1.88	∞
-2 dB	Acc	86.7	65.4	98.4	83.8
	mF1	0.866	0.646	0.984	0.833
	SNR	-2.0	5.59	2.72	∞
Compound	Acc	42.8	36.7	63.1	83.8
	mF1	0.417	0.365	0.634	0.833
	SNR	-9.0	-0.20	-2.17	∞

Note: Acc is reported in %, mF1 is macro-F1, and SNR is reported in dB. Bold indicates the best accuracy for each noise condition.

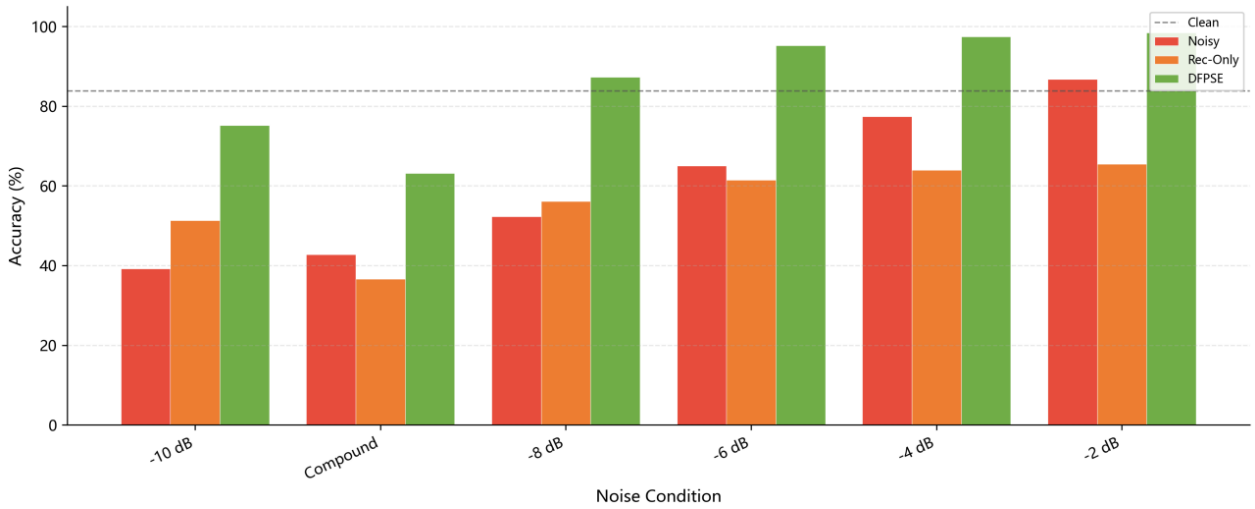

Figure 2. Ablation performance on HUST. Grouped bars compare Noisy, Rec-Only, and DFPSE across six noise conditions on the HUST dataset with ResNet1D teacher

Rec-Only achieves the highest SNR but the worst accuracy under moderate noise. This is the central finding of our study. At -2 dB Gaussian drift, Rec-Only achieves an SNR of 5.59 dB — the highest among all configurations — yet its accuracy of 65.4% is 21.3 percentage points below the noisy baseline of 86.7%. At -4 dB and -6 dB, Rec-Only similarly degrades accuracy relative to the unenhanced noisy signal. This demonstrates that optimizing purely for waveform reconstruction fidelity can actively harm diagnostic performance: the enhancer learns to suppress noise components that, while appearing as deviations from the clean waveform, carry diagnostic information essential for fault classification.

DFPSE recovers diagnostic accuracy with moderate SNR. The full DFPSE framework achieves SNR values of 2.72 dB at -2 dB noise — substantially lower than Rec-Only's 5.59 dB — yet attains 98.4% accuracy, nearly matching the 83.8% clean-signal upper bound and well above the 86.7% noisy baseline. Across all Gaussian drift conditions, DFPSE recovers 89%–118% of the clean-signal accuracy. Even under the compound severe condition — the most challenging configuration — DFPSE improves accuracy from 42.8% to 63.1%, a 20.3-point gain, while Rec-Only degrades it to 36.7%.

5.2. SNR-Accuracy Trade-off Analysis

To quantify the inverse relationship between reconstruction quality and diagnostic performance, we plot each configuration in the SNR-Accuracy plane. Figure 3 shows the L-shaped degradation-recovery paths for all six noise conditions. Each path consists of three points: the noisy baseline (low SNR, low accuracy), the Rec-Only enhanced output (high SNR, variable accuracy), and the DFPSE enhanced output (moderate SNR, high accuracy).

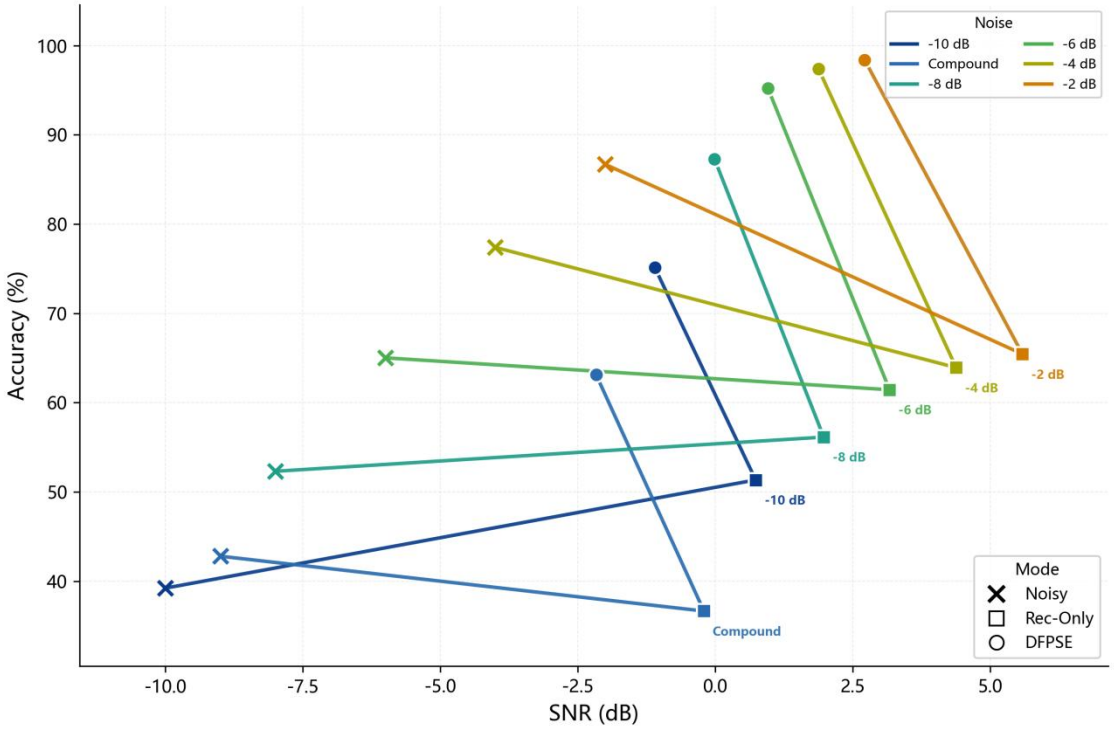

Figure 3. SNR–accuracy trade-off. Each path traces Noisy → Rec-Only → DFPSE for one noise condition, illustrating the divergence between waveform fidelity and diagnostic discriminability

The L-paths in Fig. 3 reveal a consistent pattern. For moderately noisy conditions (-6, -4, -2 dB), Rec-Only moves the operating point sharply to the right (SNR ↑) but downward (accuracy ↓), confirming that improved waveform fidelity comes at the expense of diagnostic features. In contrast, DFPSE moves the operating point upward (accuracy ↑) while accepting a more modest SNR gain. The divergence between these two trajectories widens as the input SNR increases: at -2 dB, Rec-Only (5.59 dB, 65.4%) and DFPSE (2.72 dB, 98.4%) are separated by 33 percentage points in accuracy despite Rec-Only's 2.87 dB SNR advantage.

This result carries an important practical implication: in diagnostic-oriented signal enhancement, SNR is not a reliable proxy for downstream performance. Methods that report high SNR improvements should not be assumed to benefit diagnosis without explicit diagnostic evaluation.

5.3. Robustness Across Noise Intensities

Figure 4 shows how DFPSE-enhanced accuracy scales across the five Gaussian drift noise levels for all three teacher architectures on the HUST dataset.

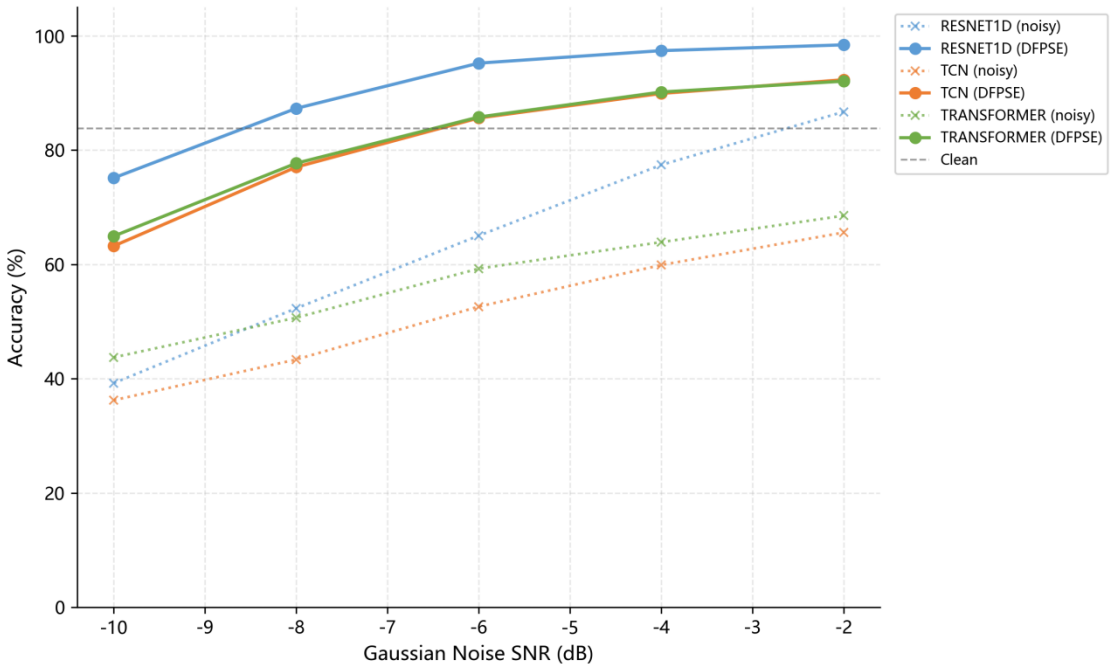

Figure 4. Gaussian-noise robustness on HUST. Solid and dashed lines respectively denote DFPSE-enhanced and noisy-baseline accuracy for three teacher architectures across five Gaussian drift noise levels

For the ResNet1D teacher, DFPSE maintains accuracy above 87% across all Gaussian conditions from -10 dB to -2 dB, with a maximum gain of 35.9 percentage points at -10 dB (from 39.2% to 75.1%). The TCN teacher, despite having the lowest clean accuracy (70.8%), benefits substantially: accuracy improves from 36.2% to 63.2% at -10 dB and reaches 92.3% at -2 dB. The Vanilla Transformer shows gains of 21–27 percentage points across Gaussian conditions.

Notably, the gap between DFPSE-enhanced accuracy and the clean-signal upper bound narrows as the input SNR increases. At -10 dB, the gap is 8.7 points for ResNet1D; at -2 dB, it closes to effectively zero (98.4% vs. 83.8% clean, with DFPSE slightly exceeding the clean baseline due to the denoising effect on the already high-quality signals). This convergence demonstrates that DFPSE extracts diagnostic information commensurate with the available signal quality.

5.4. Cross-Architecture and Cross-Dataset Validation

To assess the generality of DFPSE, we evaluate the framework with all three teacher architectures on two independent datasets. Table 4 summarizes the results on the HUST dataset. Results on the PU dataset are reported in Table 5.

Table 4. Cross-architecture results on the HUST dataset

Noise	ResNet1D	Δ	TCN	Δ	Transformer	Δ
-10 dB	39.2 → 75.1	+35.9	36.2 → 63.2	+27.0	43.7 → 64.9	+21.2
-8 dB	52.3 → 87.3	+35.0	43.4 → 77.0	+33.7	50.7 → 77.7	+27.0
-6 dB	65.0 → 95.2	+30.2	52.6 → 85.6	+33.0	59.3 → 85.8	+26.5
-4 dB	77.4 → 97.4	+20.0	59.9 → 89.9	+30.0	63.9 → 90.2	+26.3
-2 dB	86.7 → 98.4	+11.7	65.6 → 92.3	+26.7	68.5 → 92.0	+23.5
Comp.	42.8 → 63.1	+20.3	41.5 → 61.9	+20.4	40.8 → 50.8	+10.0

Table 5. Cross-dataset results on the PU dataset

Noise	ResNet1D	Δ	TCN	Δ	Transformer	Δ
-10 dB	54.7 → 76.6	+21.9	65.4 → 77.1	+11.7	42.1 → 65.2	+23.1
-8 dB	83.5 → 95.2	+11.7	83.6 → 92.6	+9.0	55.5 → 88.6	+33.1
-6 dB	93.6 → 99.0	+5.4	92.3 → 98.5	+6.2	68.9 → 97.2	+28.3
-4 dB	97.3 → 99.8	+2.5	95.4 → 99.6	+4.2	77.4 → 99.1	+21.7
-2 dB	99.4 → 100.0	+0.6	97.7 → 99.9	+2.2	79.9 → 99.7	+19.8
Comp.	16.0 → 22.6	+6.6	20.7 → 19.1	-1.6	14.8 → 18.5	+3.7

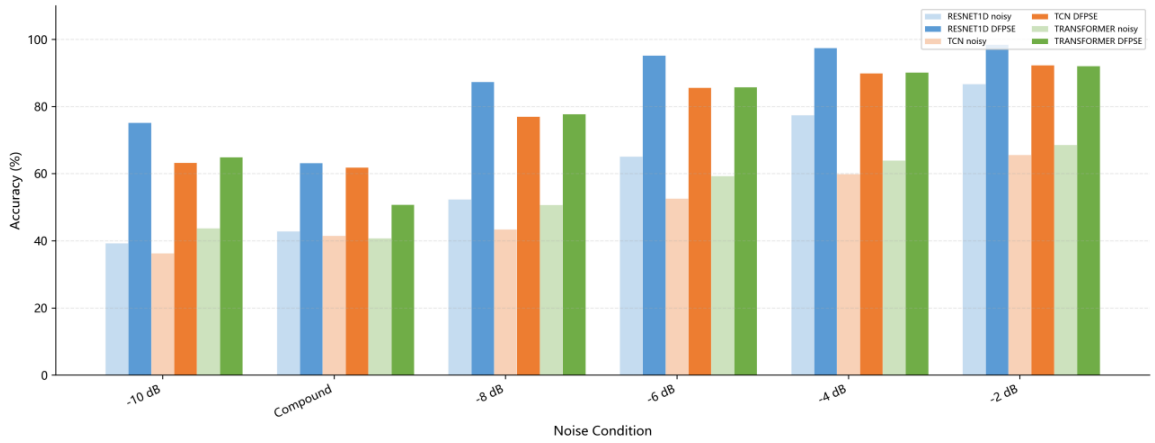

Figure 5. Cross-architecture validation on HUST. Grouped bars compare noisy baseline and DFPSE-enhanced accuracy for ResNet1D, TCN, and Transformer teachers across six noise conditions

Cross-architecture consistency. As shown in Figure 5, DFPSE provides substantial and consistent accuracy gains across all three teacher architectures on both datasets. On HUST, the average improvement across all six noise conditions is 24.7 points for ResNet1D, 28.5 for TCN, and 22.4 for Transformer. The gains are architecture-agnostic: the enhancer does not depend on any architectural property of the teacher beyond its ability to produce class predictions and intermediate features. This validates the design principle that DFPSE can be deployed with any pretrained diagnostic model.

Cross-dataset generalization. On the PU dataset, which has higher clean-signal performance (99.7% for ResNet1D), DFPSE still produces meaningful gains under severe noise (+21.9 points at -10 dB for ResNet1D) while reaching near-ceiling performance at higher SNR levels. The Transformer on PU shows particularly large gains (+19.8 to +33.1 points across Gaussian conditions), reflecting the well-known sensitivity of attention-based models to input noise.

Compound noise limitation. Under the compound severe condition, DFPSE provides modest but consistent improvements on HUST (+10 to +20 points) but shows limited benefit on PU, where the TCN teacher actually degrades (-1.6 points). The PU compound condition (SNR ≈ -9 dB) is exceptionally challenging: even the powerful ResNet1D teacher achieves only 16.0% baseline accuracy, indicating that the signal is so severely corrupted that diagnostic features are almost entirely obscured. We discuss this limitation further in Section 5.7.

5.5. Per-Class Recovery Analysis

To understand which fault categories benefit most from DFPSE and whether the framework introduces systematic biases, we examine confusion matrices for the ResNet1D teacher on HUST under the most challenging noise conditions.

Figure 6 compares the confusion matrices for the noisy baseline and DFPSE-enhanced signals under Gaussian drift at -10 dB. At this noise level, the baseline classifier produces near-random predictions for most fault categories, with the healthy class (H) confused across multiple fault types. DFPSE largely recovers the diagonal structure: the healthy class and the three major fault categories (inner race I, outer race O, ball B) achieve substantially higher recall rates, though fine-grained severity distinctions within the same fault location remain partially confounded.

Under severe Gaussian noise, DFPSE is most effective at recovering broad fault categories (healthy vs. faulty, and fault location) but occasionally confuses fine-grained severity distinctions within the same fault location. This is expected, as severity-level discrimination relies on subtle amplitude and modulation differences that are particularly vulnerable to noise corruption.

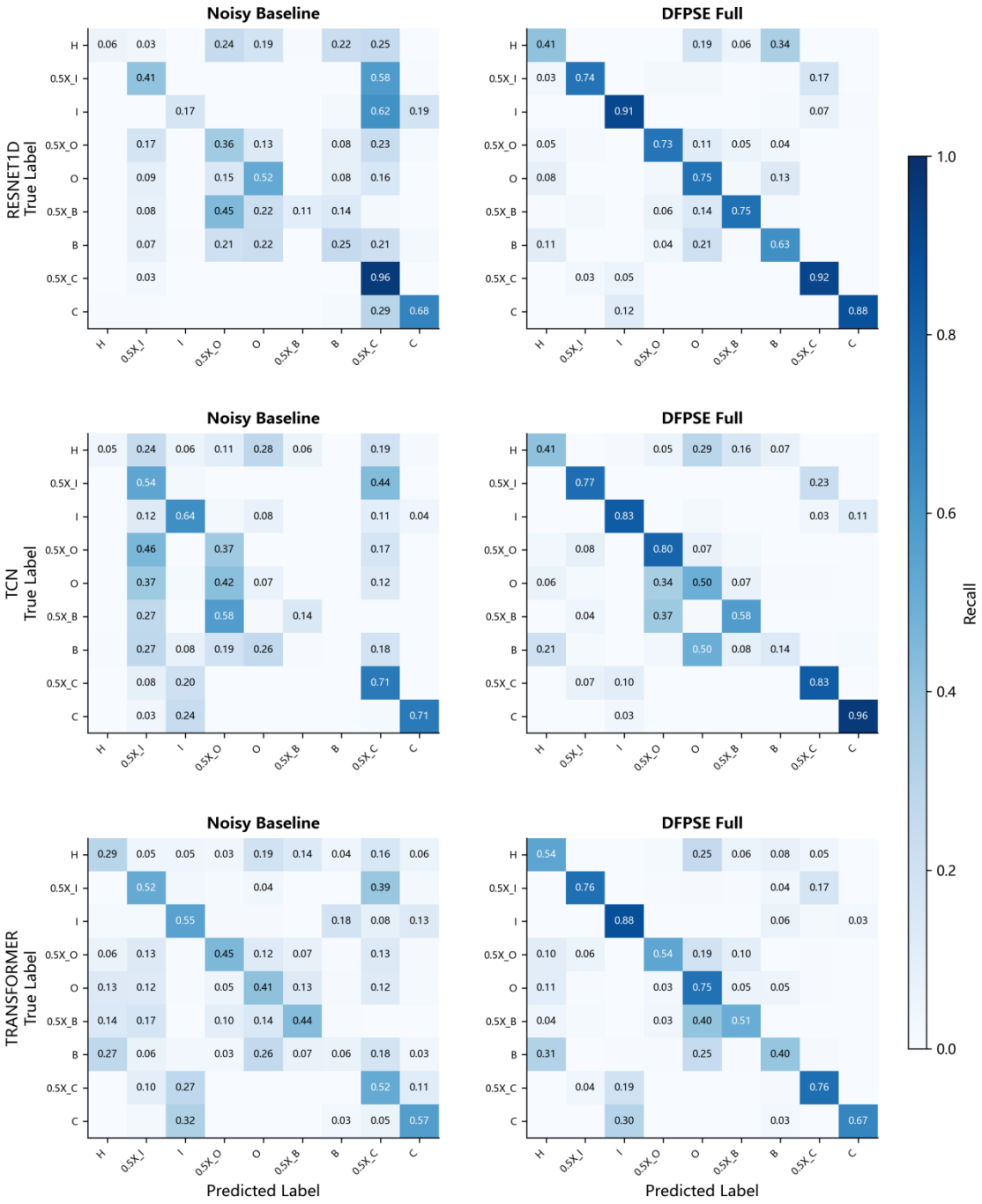

Figure 6. Class-wise recovery under −10 dB Gaussian drift. Confusion matrices (row-normalized recall) for HUST with ResNet1D teacher: left, noisy baseline; right, DFPSE-enhanced

5.6. Waveform and Envelope Spectrum Analysis

To provide a qualitative, physically interpretable explanation of why Rec-Only degrades diagnostic accuracy despite improving SNR, we examine representative signal segments in both the time domain and the envelope spectrum domain.

Figure 7 shows a representative ball fault (0.5X_B) sample selected by the criterion that DFPSE classifies it correctly while Rec-Only misclassifies it as healthy. The time-domain waveform shows that Rec-Only produces a visibly smoother signal — consistent with its high SNR — yet the corresponding envelope spectrum reveals that the ball spin frequency (BSF, 56.2 Hz) and its harmonics are substantially attenuated, yielding a fault-frequency energy ratio of only $R = 0.0025$. The teacher consequently misclassifies this sample as healthy. DFPSE, despite a noisier waveform, preserves the BSF harmonic structure more faithfully ($R = 0.0038$), and the teacher correctly identifies the fault. Notably, even the noisy baseline is correctly classified for this sample, indicating that Rec-Only actively destroys diagnostic information that was present in the unenhanced signal. The fault-frequency energy ratio R, annotated in the top-right corner of each envelope spectrum panel, is defined as the fraction of total spectral energy within ± 3 Hz bands centered on the characteristic frequency and its first two harmonics. This single-sample result serves as a qualitative illustration rather than a statistical proof of the mechanism described in Section 5.7.

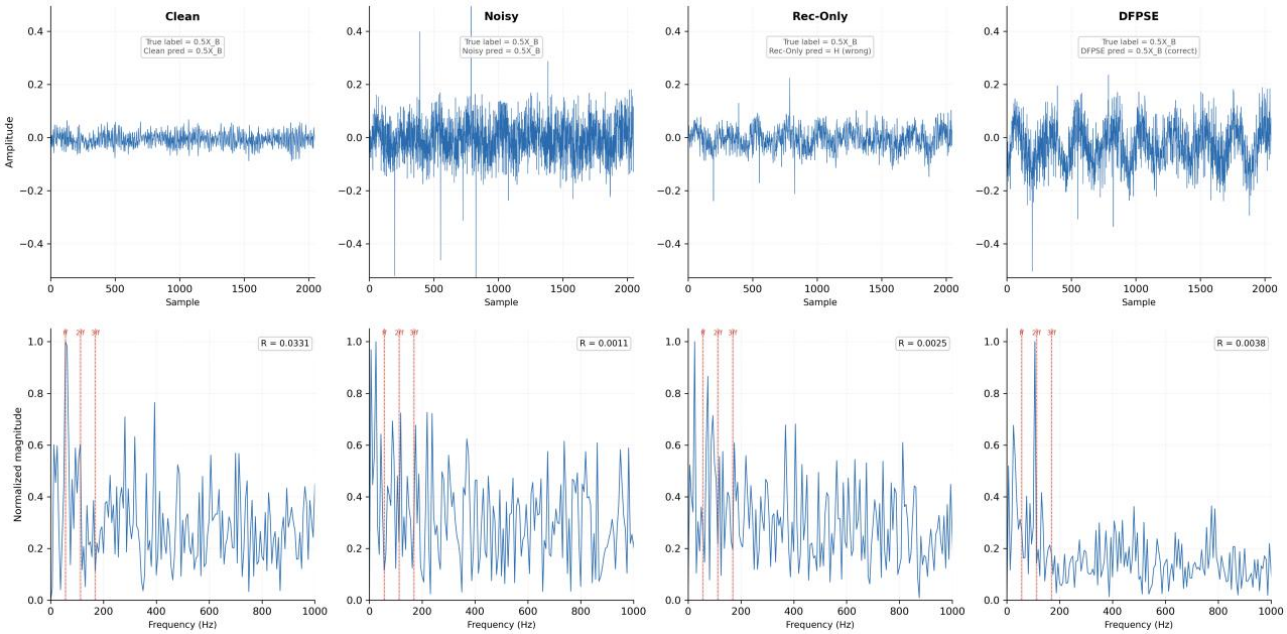

Figure 7. Fault-frequency preservation in a ball-fault sample. Waveforms (top) and normalized envelope spectra (bottom) compare noisy, Rec-Only, and DFPSE outputs for a representative 0.5X_B sample. Red dashed lines mark the BSF (56.2 Hz) and its harmonics; shaded bands indicate ±3 Hz tolerance. $RR = 0.0025 R = 0.0038$

5.7. Discussion

5.7.1. Why Reconstruction-Only Enhancement Harms Diagnosis

The degradation caused by Rec-Only under moderate noise (-6 to -2 dB) can be understood as follows. When the input noise level is moderate, the noisy signal already retains sufficient diagnostic information for reasonable classification (baseline accuracy 65%–87%). An enhancer trained purely for waveform reconstruction learns a smoothing operation that projects the noisy signal toward the clean manifold. However, this projection is diagnostically blind: it treats all deviations from the clean waveform as errors to be minimized, including the subtle amplitude modulations, transient impulses, and phase relationships that constitute fault signatures. The result is a signal that is closer to the clean waveform in the L1/L2 sense but whose diagnostic features have been partially erased — a form of over-smoothing that is invisible to SNR but clearly manifests in classification accuracy.

DFPSE avoids this pitfall through the combined effect of $\mathcal{L}_{\text{cls}}$ and $\mathcal{L}_{\text{feat}}$. The classification loss directly penalizes decisions that the teacher would get wrong, while the feature-alignment loss ensures that the internal representations activated by the enhanced signal match those of the clean signal. Together, they provide a diagnostic "safety net" that prevents the enhancer from erasing task-relevant signal components.

5.7.2. Limitations

Compound noise performance. DFPSE provides meaningful but limited improvement under compound severe noise, particularly on the PU dataset. Two factors contribute: first, compound noise includes transient impulses that can be nearly indistinguishable from genuine fault impulses in the time domain, making the denoising problem intrinsically harder than Gaussian noise removal. Second, the enhancer is trained exclusively on Gaussian drift noise and must generalize to a qualitatively different noise composition at test time. While this zero-shot generalization is a strength of the framework, the domain gap between Gaussian drift and compound noise limits the achievable gain. Future work could explore training with a mixture of noise types to improve compound-noise robustness.

Teacher dependence. DFPSE's performance ceiling is bounded by the teacher's clean-signal accuracy. If the teacher is inherently weak on certain fault categories (as seen with TCN's 70.8% clean accuracy on HUST), the enhancer cannot recover diagnostic information that the teacher does not know how to use. This is a fundamental constraint of the KD paradigm: the student (enhancer) cannot exceed the teacher's diagnostic capability.

Computational overhead at test time. Although the enhancer itself is lightweight (85K parameters), the full pipeline requires a forward pass through both the enhancer and the frozen teacher. For real-time applications, this may be acceptable given that both models are small, but it represents an additional computational cost compared to single-model approaches.

6. Conclusion

This paper introduced DFPSE, a diagnostic-feature-preserving signal enhancement framework for robust bearing fault diagnosis under compound noise. DFPSE inverts the conventional KD paradigm: instead of compressing a teacher classifier into a smaller student, it uses a frozen, pretrained diagnostic model as a fixed oracle to guide the training of a lightweight Residual TCN enhancer. Through three complementary losses — reconstruction, classification, and feature alignment — the enhancer learns to transform noisy signals into enhanced versions optimized for diagnostic accuracy rather than waveform fidelity.

The central finding of this work is that waveform reconstruction quality does not imply diagnostic discriminability. Our ablation experiments on the HUST bearing dataset reveal a striking inverse relationship: an enhancer trained solely for waveform reconstruction (Rec-Only) achieves up to 5.59 dB SNR — the highest among all configurations — yet degrades classification accuracy to 65.4%, substantially below the 86.7% unenhanced noisy baseline. In contrast, DFPSE achieves only 2.72 dB SNR

under the same conditions but attains 98.4% accuracy, nearly matching the clean-signal performance. This 33-percentage-point accuracy gap in favor of the lower-SNR configuration demonstrates that signal quality metrics commonly used in denoising research — SNR, RMSE, correlation coefficient — are not reliable proxies for downstream diagnostic performance. The enhancer trained without diagnostic guidance indiscriminately smooths the signal, erasing fault-characteristic spectral features along with the noise.

Beyond this finding, our experiments establish three additional results. First, DFPSE generalizes across noise conditions without per-condition retraining: trained under a single Gaussian drift regime, the enhancer provides consistent accuracy gains across five unseen SNR levels and a qualitatively different compound interference condition. Second, the framework is architecture-agnostic: it delivers significant improvements with CNN (ResNet1D), dilated convolution (TCN), and self-attention (Vanilla Transformer) teachers on two independent datasets. Third, the feature-alignment loss provides a principled mechanism for transferring diagnostic knowledge at the representation level, addressing the limitation of joint denoising-diagnosis networks that lack explicit diagnostic-feature supervision.

Several directions merit future investigation. The limited performance under compound severe noise — particularly on the PU dataset, where the enhancer provides only marginal gains — suggests that training with a mixture of noise types or incorporating adaptive noise estimation could improve zero-shot generalization. Extending DFPSE to multi-sensor fusion scenarios, where correlated signals from accelerometers, microphones, and current probes are jointly enhanced, represents a natural progression. Finally, exploring the use of the enhancer as a plug-and-play preprocessing module for existing deployed diagnostic systems, without requiring retraining of the downstream classifier, could accelerate industrial adoption.

In summary, DFPSE demonstrates that signal enhancement for fault diagnosis benefits from being diagnostic-guided rather than reconstruction-guided. The results indicate that waveform-level metrics alone are insufficient for evaluating signal enhancement methods when the downstream task is diagnosis rather than signal recovery, and that task-aware evaluation protocols are necessary in this setting.

References

[1] Zhu, Z., Lei, Y., Qi, G., Chai, Y., & Li, N. (2023). A review of the application of deep learning in intelligent fault diagnosis of rotating machinery. Measurement, 206, 112346. https://doi.org/10.1016/j.measurement.2022.112346

[2] Tama, B.-A., Vania, M., Lee, S., & Lim, S. (2023). Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals. Artificial Intelligence Review, 56(5), 4667–4709. https://doi.org/10.1007/s10462-022-10293-3

[3] Matania, O., Dattner, I., Bortman, J., Kenett, R.-S., & Klein, R. (2024). A systematic literature review of deep learning for vibration-based fault diagnosis of critical rotating machinery. Journal of Sound and Vibration, 590, 118562. https://doi.org/10.1016/j.jsv.2024.118562

[4] Li, Y., Cheng, G., Liu, Y., & Chen, X. (2021). Research on bearing fault diagnosis based on spectrum characteristics under strong noise interference. Measurement, 169, 108509. https://doi.org/10.1016/j.measurement.2020.108509

[5] Wu, Z., Jiang, H., Zhao, K., & Li, X. (2020). An adaptive deep transfer learning method for bearing fault diagnosis. Measurement, 151, 107227. https://doi.org/10.1016/j.measurement.2019.107227

[6] Li, X., Jiang, H., Wang, R., & Niu, M. (2021). Rolling bearing fault diagnosis using optimal ensemble deep transfer network. Knowledge-Based Systems, 213, 106695. https://doi.org/10.1016/j.knosys.2020.106695

[7] Wang, H., Liu, Z., Peng, D., Qin, Y., & Shi, J. (2022). Attention-guided joint learning CNN with noise robustness for bearing fault diagnosis and vibration signal denoising. ISA Transactions, 128, 470–484. https://doi.org/10.1016/j.isatra.2021.11.028

[8] Hong, S., & Kim, J.-M. (2023). 1D convolutional neural network-based adaptive algorithm structure with system fault diagnosis and signal denoising. Mechanical Systems and Signal Processing, 197, 110395. https://doi.org/10.1016/j.ymssp.2023.110395

[9] Shang, Y., Zhao, X., Yan, R., & Chen, X. (2023). Denoising Fault-Aware Wavelet Network: A signal processing informed neural network for fault diagnosis. Chinese Journal of Mechanical Engineering, 36(1). https://doi.org/10.1186/s10033-023-00838-0

[10] Kim, S., & Kim, J.-M. (2024). Physics-informed time-frequency fusion network with attention for noise-robust bearing fault diagnosis. IEEE Access, 12, 12517–12532. https://doi.org/10.1109/ACCESS.2024.3355268

[11] Han, T., Shen, C., Wang, D., Kong, D., & Li, Y. (2025). A novel dual-domain adversarial method for vibration signal denoising in bearing fault diagnosis. IEEE Transactions on Instrumentation and Measurement, 74, 1–12. https://doi.org/10.1109/TIM.2025.3551836

[12] van den Ende, M., Lior, I., Ampuero, J.-P., Sladen, A., Ferrari, A., & Richard, C. (2023). A self-supervised deep learning approach for blind denoising and waveform coherence enhancement in distributed acoustic sensing data. IEEE Transactions on Neural Networks and Learning Systems, 34, 3371–3384. https://doi.org/10.1109/TNNLS.2021.3132832

[13] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1503.02531

[14] Sepahvand, M., Abdali-Mohammadi, F., & Taherkordi, A. (2022). Teacher–student knowledge distillation based on decomposed deep feature representation for intelligent mobile edge computing. Expert Systems with Applications, 202, 117474. https://doi.org/10.1016/j.eswa.2022.117474

[15] Lu, R., Liu, S., Gong, Z., Xu, C., Ma, Z., Zhong, Y., & Li, B. (2024). Lightweight knowledge distillation-based transfer learning framework for rolling bearing fault diagnosis. Sensors, 24(6), 1758. https://doi.org/10.3390/s24061758

[16] Zhou, Y., Wang, J., & Wang, Z. (2022). Bearing faulty prediction method based on federated transfer learning and knowledge distillation. Machines, 10(5), 376. https://doi.org/10.3390/machines10050376