Computers and Artificial Intelligence；Vol.3, No.1; 2026
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.26v3n1.0116
Published by: Zeus Press

HAGAN: A Lightweight Compressed Sensing Framework for Motor Bearing Fault Diagnosis

Weibing Tang[1,*], Hang Liu[1]

[1]Dept. of Student, Harbin University of Science and Technology, Harbin, Heilongjiang, China

**Corresponding author: Weibing Tang*

Abstract

Motor bearings are the most failure-prone components in industrial motors, and the accuracy of their fault diagnosis is highly dependent on the effective acquisition and analysis of vibration signals. However, traditional Compressed Sensing (CS) methods face an inherent trade-off between compression ratio and reconstruction accuracy, while deep learning-enhanced CS models generally suffer from complex architectures and insufficient real-time performance. These drawbacks severely restrict their practical application in industrial fault diagnosis scenarios. To address the aforementioned challenges, this study proposes a lightweight compressed sensing framework tailored for motor bearing fault diagnosis-Hybrid Autoencoder Generative Adversarial Network (HAGAN). This framework integrates the core advantages of Autoencoders (AE) and Generative Adversarial Networks (GAN). The AE is responsible for extracting key fault features and compressing high-dimensional vibration signals, while the GAN improves the fidelity of reconstructed signals through an adversarial training mechanism. Meanwhile, a streamlined network structure design is adopted to remove redundant nonlinear modules, thereby reducing computational overhead and ensuring real-time deployment capabilities in industrial settings. In the experiments, this method achieves an ultra-high data compression ratio of 100:1 for motor bearing fault data. Three metrics-Root Mean Square Error (RMSE), Percentage Root Mean Square Deviation (PRD), and Signal-to-Noise Ratio (SNR)-are employed to comprehensively verify the compression and reconstruction performance. The results demonstrate that the HAGAN framework can effectively preserve the characteristic information of bearings in both healthy states and various fault states (including inner race faults, outer race faults, rolling element faults, and combination faults) even under high compression ratios. Its compression and reconstruction performance, quantified and validated by the three metrics, is excellent. The framework can be successfully applied to data compression and reconstruction tasks in motor bearing fault diagnosis, providing an efficient and reliable fault diagnosis solution for resource-constrained industrial environments.

Keywords

motor bearing fault diagnosis, compressed sensing, HAGAN, autoencoder, generative adversarial network (GAN), signal compression and reconstruction

1. Introduction

Motors serve as the most ubiquitous industrial electric drives in modern manufacturing, underpinning a vast array of processes and applications [1]. The effective identification of motor faults is a critical prerequisite for ensuring safe and uninterrupted production. Proactive fault diagnosis can prevent severe economic consequences, including equipment damage and production downtime, which often result from unforeseen failures [2]. To address this challenge, significant research and engineering efforts have been devoted to early-stage fault diagnosis and prediction. Statistics indicate that bearing-related defects constitute approximately 45%-55% of all mechanical failures in motors, establishing bearings as the most failure-prone component [3].

Conventional sampling frameworks are primarily dominated by the Nyquist-Shannon theorem and the more recent compressed sensing (CS) paradigm. The Nyquist-Shannon sampling theorem serves as the cornerstone of traditional signal acquisition. It mandates that to perfectly reconstruct a bandlimited signal without aliasing, the sampling rate must be at least twice the signal's highest frequency component (known as the Nyquist rate) [4]. Consequently, the high sampling rates required for high-frequency signals result in substantial data volumes, imposing significant burdens on storage and transmission systems. If a signal is sampled below the Nyquist rate, aliasing occurs, causing the spectral replicates to overlap. Attempting to mitigate aliasing solely through low-pass filtering after sub-Nyquist sampling inevitably distorts the signal's frequency content within the overlapping regions, making perfect recovery of the original time-domain signal impossible [5]. Moreover, traditional data compression approaches typically operate by first acquiring the full Nyquist-rate dataset and then employing compression algorithms to remove redundancy. This *sample-then-compress* paradigm necessitates extensive computational resources, transmission bandwidth, and storage capacity[6]. Compressed sensing (CS) is an innovative signal processing framework that simultaneously performs sampling and data compression. It fundamentally bypasses the strict requirements of the Nyquist-Shannon theorem by leveraging signal sparsity, enabling the faithful reconstruction of signals from a number of measurements significantly lower than the Nyquist rate [7]. This capability offers significant potential for efficient data acquisition in resource-constrained environments, including wireless sensor networks, biomedical imaging, and remote sensing systems.

In recent years, machine learning has undergone rapid advancement, driven by continuous technological innovations. It has overcome limitations inherent in traditional techniques across diverse fields such as image processing [8], machine vision, and speech recognition [9], thereby introducing novel paradigms for motor bearing fault diagnosis. The distinct characteristic frequencies exhibited by motor bearings under healthy conditions and various fault states provide a foundational basis for fault diagnosis. The recent proliferation of deep learning and other data-driven models has demonstrated superior performance compared to traditional methods. CS reconstruction algorithms, including matching pursuit, iterative hard thresholding (IHT), and Bayesian methods, have proven effective at recovering sparse signals from randomly acquired, incomplete measurements [6]. However, when dealing with high-dimensional or complex signals, the reconstruction performance of traditional CS methods is often sensitive to the number of measurements (compression ratio). An excessive number of measurements diminishes the benefits of compression, whereas an insufficient number compromises reconstruction accuracy.

Motivated by the success of deep learning in complex tasks, this work proposes a novel framework that integrates autoencoders (AEs), generative adversarial networks (GANs), and traditional CS principles. The objective is to overcome the inherent trade-off between compression ratio and reconstruction accuracy prevalent in conventional CS methods. Our approach aims to significantly reduce data volume while ensuring the faithful preservation of critical information for accurate fault diagnosis. Specifically, by streamlining the encoder network architecture through the removal of non-linear modules and the strategic batching of normalization and activation layers, our model facilitates high-quality signal reconstruction with enhanced computational efficiency and reduced inference time [10].

The aforementioned challenges inherent in traditional CS methods, particularly when dealing with complex mechanical signals, have motivated the exploration of deep learning-enhanced CS frameworks. These hybrid approaches seek to leverage the representational power of neural networks to learn optimal signal representations and reconstruction mappings from data, potentially overcoming the limitations of hand-crafted sparsity bases and linear reconstruction algorithms. For instance, Kulkarni et al. [11] were among the first to employ deep neural networks for reconstructing images from random measurements, pioneering the concept

of learned compressed sensing reconstruction. Subsequent approaches, such as ReconNet [11] and DeepInverse [12], further demonstrated that neural network architectures can significantly outperform traditional iterative algorithms in terms of both reconstruction accuracy and computational efficiency. In recent years, model-driven deep unfolding networks [13] and learnable sensing matrix optimization methods [14] have bridged the gap between theoretical CS models and data-driven learning frameworks. Therefore, incorporating deep neural networks into the compressed sensing paradigm has become a promising direction to overcome the inherent limitations of conventional model-driven approaches.

However, most existing deep compressed sensing (CS) models inherit considerable architectural complexity from generic deep architectures. This complexity introduces substantial computational overhead during both the encoding and reconstruction phases, thereby undermining the benefits of data reduction for real-time industrial fault diagnosis applications. For example, methods based on deep convolutional networks [11] or autoencoder-like structures [15] often retain the heavy network design of conventional deep architectures, compromising the real-time performance of compressed sensing in industrial environments. Furthermore, effectively training such complex models under extreme compression ratios remains a non-trivial challenge. When the measurement rate is too low, the reconstruction network's ability to recover discriminative features often deteriorates, resulting in poor reconstruction fidelity and, consequently, degraded accuracy in subsequent fault classification and anomaly detection [16].

To address these limitations, we propose a novel lightweight compressed sensing framework tailored for motor bearing fault diagnosis. Our main contributions are summarized as follows:

1. Lightweight Network Architecture: We propose a streamlined deep compressed sensing (CS) architecture by removing redundant nonlinear modules and unnecessary network layers. This design significantly reduces computational overhead and memory consumption, enabling real-time deployment in industrial fault diagnosis systems.

2. GAN-Based Integrated Reconstruction: We introduce a generative adversarial network (GAN) into the reconstruction process to enhance feature recovery. This approach improves reconstruction fidelity and preserves subtle fault features, ensuring that critical diagnostic information remains intact even under extremely high compression ratios.

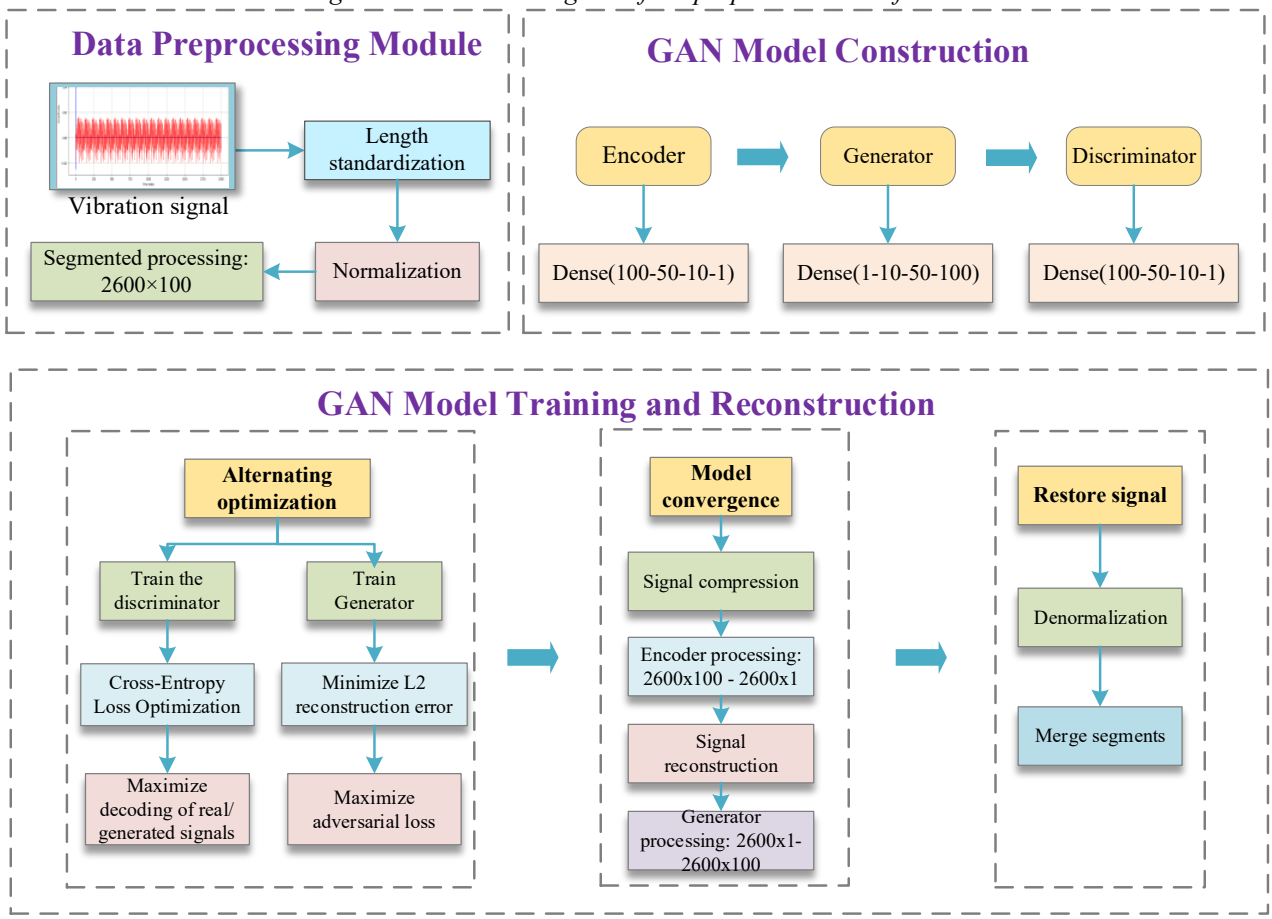
3. Comprehensive Experimental Validation: We conduct extensive experiments on benchmark bearing datasets. The results demonstrate that the proposed framework outperforms both traditional CS methods and state-of-the-art deep learning approaches in terms of reconstruction quality and downstream fault classification accuracy, confirming its effectiveness and practical applicability.

The rest of the paper is organized as follows. Section II reviews related work on compressed sensing and deep learning for fault diagnosis. Section III details the architecture and algorithm of proposed method. Section IV describes the experimental setup, presents the results, and discuss them. Finally, Section V concludes the paper and outlines future research directions.

2. Proposed Method

To effectively retain the critical diagnostic characteristics of motor bearing vibration signals during the processes of compression and reconstruction, we propose a Hybrid Autoencoder Generative Adversarial Network (HAGAN). This framework integrates the representational capability of autoencoders with the generative refinement of adversarial learning. The following section elaborates on the overall architecture of HAGAN, its training methodology, and essential implementation strategies, providing an in-depth overview of the model's structure and its superior performance in bearing fault diagnosis.

Figure 1: Schematic diagram of the proposed HAGAN framework.



2.1 Signal Compression Based on Autoencoder

Deep learning models called autoencoders are frequently employed for data compression and dimensionality reduction. They are composed of two components: a decoder that transfers information from the hidden space back to the high-dimensional data, and an encoder that compresses high-dimensional sequential signals and extracts representations in a latent feature space. We can compress data while maintaining fault features by training autoencoders.

We first preprocess the data by mapping the raw vibration signals $x_{\text{raw}}[i]$ to the $[0,1]$ interval using equation (1), thereby eliminating variations in vibration amplitude across different loads and rotational speeds.

$$x_{\text{norm}}[i] = \frac{x_{\text{raw}}[i] - x_{\min}}{x_{\max} - x_{\min}}, \qquad i = 1, \ldots, L \tag{1}$$

where $x_{\text{raw}}[i]$ signifies the original (non-normalized) i-th time-series sample, $x_{\min} = \min_i x_{\text{raw}}[i]$, $x_{\max} = \max_i x_{raw}[i]$ and $x_{\text{norm}}[i]$ indicates the normalized i-th sample.

To capture the transient impact characteristics of bearing faults, we decompose the continuous vibration signal into localized segments, specifically utilizing equation (2) to partition the normalized long sequence into S segments, each of length n, thereby forming matrix $\mathbf{X} \in$ $^{S\times n}$.

$$\mathbf{X} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(S)} \end{bmatrix} \in \quad ^{S\times n} \tag{2}$$

$$x^{(s)} = [x_{\text{norm}}[(s-1)n+1], \ldots, x_{\text{norm}}[sn]]$$

where X is a stacked matrix representing a dataset of size $S \times n$. Each data segment's length is denoted by n, and the number of segments is denoted by S, meaning that $S = L / n$.

The encoder compresses high-dimensional vibration segments into low-dimensional latent vectors. That is, the input segmented vibration data $x \in$ ^{n} is gradually compressed through two ReLU() activated hidden layers. While compressing the data volume, the key discriminative features of bearing faults are preserved.

The first layer activation:

$$h^{(1)} = \varphi_1\left(W^{(1)}x + b^{(1)}\right) \in \quad ^{50} \tag{3}$$

The second layer activation:

$$h^{(2)} = \varphi_2\left(W^{(2)}h^{(1)} + b^{(2)}\right) \in \quad ^{10} \tag{4}$$

The bottleneck layer:

$$z = E(x) = W^{(3)}h^{(2)} + b^{(3)} \in \quad ^{m} \tag{5}$$

where $\varphi_1(\cdot) = \mathrm{ReLU}(\cdot)$ and $\varphi_2(\cdot) = \mathrm{ReLU}(\cdot)$. $E(\cdot)$ denotes the encoder network mapping, which projects the input x to its reconstructed representation $\hat{x}$. The hidden activation vectors corresponding to the first, second, and bottleneck layers of the encoder are denoted as $h^{(1)}$, $h^{(2)}$, and $h^{(2)}$, respectively. The weight matrices of these layers are represented by $W^{(1)}$, $W^{(2)}$, and $W^{(3)}$, while the corresponding bias vectors are denoted by $b^{(1)}$, $b^{(2)}$, and $b^{(3)}$, respectively.

MSE reconstruction loss (for a sample segment):

$$\mathrm{L}_{\mathrm{MSE}}(x, \hat{x}) = \frac{1}{n}\sum_{i=1}^{n}\left(x[i] - \hat{x}[i]\right)^2 \tag{6}$$

For a batch of samples with a batch size of , the batch MSE is:

$$\mathrm{L}_{\mathrm{MSE}}^{\mathrm{batch}} = \frac{1}{B}\sum_{j=1}^{B}\mathrm{L}_{\mathrm{MSE}}(x^{(j)}, \hat{x}^{(j)}) \tag{7}$$

2.2 Signal Reconstruction Optimization Based on Generative Adversarial Network

To improve the feature fidelity and generation quality of the model at extreme compression ratios, we introduce a generative adversarial network. Generative adversarial networks (GANs) are a generative model consisting of a generator and a discriminator. The generator produces new data from the noisy space, while the discriminator is used to determine the difference in distribution between the reconstructed signal and the real signal. As the generator and discriminator compete with each other, they gradually reach an equilibrium point, and the generator is able to generate samples that are closer to real data.

In this framework, the generator acts as a decoder, playing a core role in reconstructing segments of the motor bearing vibration signal from low-dimensional latent vectors. Its structure is symmetrical to that of the encoder. It gradually expands the feature dimension through a ReLU()-activated hidden layer, and finally generates a reconstructed signal with the same length as the original segment through a linear output layer.

The first layer activation:

$$g^{(1)} = \phi_1\left(U^{(1)}z + c^{(1)}\right) \in \quad ^{10}, \qquad \phi_1 = \mathrm{Re}LU \tag{8}$$

Where $g^{(1)}$ is the hidden activation vector of the first layer of the generator, $U^{(1)}$ is the weight matrix of the first layer of the generator, and $c^{(1)}$ is the bias matrix of the first layer of the generator.

The second layer activation:

$$g^{(2)} = \phi_2\left(U^{(2)}g^{(1)} + c^{(2)}\right) \in \quad ^{50}, \quad \phi_2 = \mathrm{ReLU} \tag{9}$$

where $g^{(2)}$ is the hidden activation vector of the second layer of the generator, $U^{(2)}$ is the weight matrix of the second layer of the generator, and $c^{(2)}$ is the bias matrix of the second layer of the generator.

Output Reconstruction:

$$\hat{x} = G(z) = U^{(3)} g^{(2)} + c^{(3)} \in \quad ^{n} \tag{10}$$

where $G(\bullet)$ is the generator network mapping, which maps $z \mapsto \hat{x}$, where $\hat{x}$ is the reconstructed segment output by the generator. $U^{(3)}$ is the weight matrix of the 3rd layer of the generator, and $c^{(3)}$ is the bias matrix of the 3rd layer of the generator.

The discriminator guides the generator to fit the signal amplitude more accurately through adversarial training, thereby effectively overcoming the problem of blurred fault features caused by smoothing in simple AE reconstruction. To address the high-frequency noise interference in bearing vibration signals, LeakyReLU() activation is employed to enhance sensitivity to weak features.

The first layer activation:

$$d^{(1)} = \text{LeakyReLU}(V^{(1)} y + e^{(1)}) \in \quad ^{64} \tag{11}$$

The second layer activation:

$$d^{(2)} = \text{LeakyReLU}(V^{(2)} d^{(1)} + e^{(2)}) \in \quad ^{32} \tag{12}$$

Output probability:

$$D(y) = \sigma\left(V^{(3)} d^{(2)} + e^{(3)}\right) \in (0,1) \tag{13}$$

Where $D(\cdot)$ is the probability of returning "true" when training, the true sample label is 1 and the generated sample label is 0; $d^{(1)}$ and $d^{(2)}$ are the hidden activation vectors of the first and second layers inside the discriminator; $V^{(1)}$, $V^{(2)}$, and $V^{(3)}$ are the weight matrices of the first, second, $e^{(1)}$,and output layers of the discriminator, respectively; and $e^{(1)}$, $e^{(2)}$, and $e^{(3)}$ are the bias vectors of each layer of the discriminator.

During the adversarial training phase, we use a binary cross-entropy adversarial loss function to constrain the learning process of the generator and discriminator.

The goal of the discriminator is to distinguish between real samples $x^{(j)}$ and generated samples $\hat{x}^{(j)}$, and its loss function is defined as:

$$\mathrm{L}_D = -\frac{1}{B}\sum_{j=1}^{B}\left[\log D(x^{(j)}) + \log\left(1 - D(\hat{x}^{(j)})\right)\right] \tag{14}$$

This loss function encourages the discriminator to output a higher probability for real samples and a lower probability for generated samples, thereby improving its discriminative ability.

Correspondingly, the goal of the generator is to make the generated samples appear as real as possible, and its adversarial loss is defined as:

$$\mathrm{L}_G^{\text{adv}} = -\frac{1}{B}\sum_{j=1}^{B}\log D\left(\hat{x}^{(j)}\right) \tag{15}$$

This loss function minimizes the probability that the generated sample is judged as false, thereby guiding the generator to fit the amplitude and distribution characteristics of the real signal during the reconstruction process, alleviating the problems of excessive smoothing of output and blurred fault features in traditional autoencoders.

2.3 The Overall HAGAN Framework and Training Strategy

By combining autoencoders and generative adversarial networks, the input data can be mapped to a low-dimensional latent space through the encoding and decoding process of the autoencoder, thereby extracting the main features of the data. Subsequently, the generative adversarial network is trained through a game

between the generator and the discriminator to further improve the quality of data reconstruction in this low-dimensional space. The core idea behind combining the two is that the autoencoder is responsible for learning the key feature representations of the data, while the generative adversarial network generates high-quality samples that are closer to the real distribution through adversarial training, thereby effectively making up for the problem of over-smoothing in the simple reconstruction process.

Overall generator loss function:

$$\mathrm{L}_G = \alpha \mathrm{L}_{MSE}^{batch} + \beta \mathrm{L}_G^{adv} \tag{16}$$

where α and β are weighting coefficients, used to adjust the trade-off between reconstruction accuracy and generation realism.

In each training iteration, the discriminator learns to distinguish between real and generated samples, and the loss at each step is defined as follows:

$$\begin{aligned} \mathrm{L}_D^{\text{step}} &= \frac{1}{2}\left(\mathrm{L}_D^{\text{real}} + \mathrm{L}_D^{\text{fake}}\right) \\ \mathrm{L}_D^{real} &= -\frac{1}{B}\sum_i \log D(x^{(i)}) \\ \mathrm{L}_D^{fake} &= -\frac{1}{B}\sum_i \log(1 - D(\hat{x}^{(j)})) \end{aligned} \tag{17}$$

The training objective of the discriminator is to maximize $\log D(x)$ and minimize $\log(1 - D(\hat{x}))$, thereby distinguishing between real samples and generated samples; while the generator reduces reconstruction error and improves the statistical realism of generated samples by minimizing the formula. This dual constraint mechanism ensures coordinated optimization of the model in terms of reconstruction quality and data distribution fidelity.

After the model training is completed, the compression and reconstruction process of the entire HAGAN framework can be divided into three stages: encoding, reconstruction, and denormalization.

Encoding Stage: During this stage, preprocessed 2600×100 vibration signal segments are input into the encoder $E(\cdot)$ of the GAN model. Features are extracted from the high-dimensional time series signals and mapped to a low-dimensional latent space, achieving data compression. This process preserves the signal's key characteristics while reducing its complexity.

$$Z = E(X) = \begin{bmatrix} z^{(1)} \\ z^{(2)} \\ \vdots \\ z^{(S)} \end{bmatrix} \in \quad ^{S \times m} \tag{18}$$

Reconstruction Phase: The generator learns the mapping relationship from the latent space to the original space, enabling high-fidelity reconstruction of the input signal. The compressed feature vector obtained during the encoding phase is decoded by the generator $G(\cdot)$ to yield the reconstructed signal (2600×100). This phase relies on adversarial training between the generator and discriminator, where the generator aims to produce signals indistinguishable from the original, ensuring high-fidelity reconstruction.

$$\hat{\mathbf{x}} = G(\mathbf{Z}) = \begin{bmatrix} \hat{\mathbf{x}}^{(1)} \\ \hat{\mathbf{x}}^{(2)} \\ \vdots \\ \hat{\mathbf{x}}^{(S)} \end{bmatrix} \in \quad ^{S \times n} \tag{19}$$

Denormalization: At this stage, the reconstructed vibration signal segments undergo denormalization to restore them to their original scale. After denormalization is complete, these segments are merged together to form the complete vibration signal.

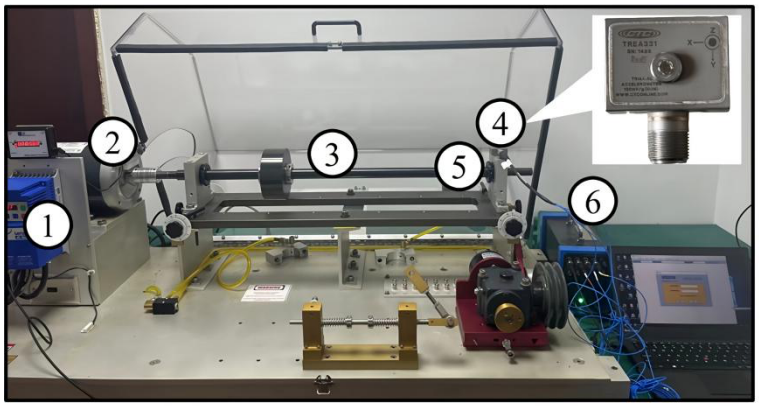
$$\hat{x}_{raw}[i] = x_{min} + \hat{x}[i] \cdot \left(x_{max} - x_{min} \right) \tag{20}$$

3. Experimental Setup

3.1 Experimental Platform

The experimental platform used for bearing vibration signal acquisition, as shown in Figure 2, consists of several key components, including a speed control system, a motor, a shaft, an acceleration sensor, a bearing, and a data acquisition board [17]. Mechanically, the system features a motor-shaft-bearing assembly, where the accelerometer is directly mounted on the bearing housing to ensure direct capture of vibration responses. During operation, the motor drives the shaft at controlled speeds, causing the bearing to generate characteristic vibrations. These vibrations are captured by the TREA331 high-precision accelerometer, which incorporates integrated signal conditioning capabilities. The resulting analog signals are then transmitted to a National Instruments NI-9234 data acquisition card, where they are sampled at 25.6 kHz for digital conversion. The digitized data is finally stored in the host computer, forming a complete signal chain that enables reliable acquisition and analysis of bearing vibration signals under various rotational conditions.

Figure 2. Experimental Platform for Bearing Vibration Signal Acquisition (adapted from [17])



1: Speed control, 2:Motor, 3: Shaft, 4: Acceleration sensor, 5: Bearing, 6: Data acquisition board

In order to collect vibration signals at different operating conditions, the experimental platform was set to different rotational speeds. As shown in Table 1, the sampling rates corresponding to these speeds were as follows:

Table 1: Sampling rates corresponding to different rotational frequency

Rotational speed (rpm)	Sampling rate (Hz)
3900	65
4200	70
4500	75
4800	80

In this experimental platform, the relevant parameters of the bearing are as shown in Table 2:

Table 2: Bearing geometric parameters

Parameter Description	Value
Nominal Diameter (D_m)	39.65 mm
Inner Diameter (D_i)	30.59 mm
Outer Diameter (D_o)	46.47 mm
Contact Angle (α)	0°
Rolling Element Diameter (D_b)	7.94 mm
Number of Rolling Elements (N_b)	9

In addition, the experimental platform can calculate the characteristic fault frequencies for different types of bearing faults based on the rotational speed f_r, as shown in Table 3.

Table 3: Characteristic fault frequencies for different bearing faults

Fault Frequency Parameter	Formula
Outer Race Fault Frequency (BPFO)	$3.591 \times \frac{f_r}{2}$
Inner Race Fault Frequency (BPFI)	$5.408 \times \frac{f_r}{2}$
Ball Fault Frequency (BSF)	$2.375 \times \frac{f_r}{2}$
Cage Fault Frequency (FTF)	$0.399 \times \frac{f_r}{2}$

3.2 Model Training Setup

In the data preprocessing step of model training, the raw signals were first truncated or zero-padded to a fixed length of L =260,000 points to ensure uniform input size. After normalization, the signals are divided into segments, each containing 100 points, resulting in a total of 2,600 segments.

In the model construction, the 100-dimensional new signal segment is fed into the Encoder $E(\cdot)$ of a three-layer fully connected network. In the first layer, the input is mapped to a 50-dimensional hidden vector $h^{(1)}$ through weights $W^{(1)}$ and biases $b^{(1)}$. This hidden vector $h^{(1)}$ is then mapped to a 10-dimensional hidden vector $h^{(2)}$ in the second layer. Finally, a linear mapping outputs a 1-dimensional latent vector Z, achieving compression of the high-dimensional signal. Input the 1-dimensional latent vector Z into the Generator $G(\cdot)$, which sequentially processes it through three fully connected layers. This process progressively reconstructs the low-dimensional signal into 10-dimensional and 50-dimensional representations, ultimately outputting a 100-dimensional reconstruction segment $\hat{x}$, thereby achieving low-to-high-dimensional reconstruction. Input the real signal segment x and the generated signal segment $\hat{x}$ into the Discriminator $D(\cdot)$. This employs a fully connected network with two layers activated by LeakyReLU(), culminating in a Sigmoid activation that outputs the authenticity probability within the (0,1) interval. This output is used to distinguish the signal source.

During the alternating training process, the autoencoder and GAN are iteratively trained to achieve optimal performance in data generation and compression tasks. Finally, the reconstructed segmented signals are merged into a complete sequence. The original signal dimensions are restored using the denormalization formula, yielding the final reconstructed vibration signal $\hat{x}_{raw}[i]$.

4. Experimental Results

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

4.1 Model Evaluation Methods

1. RMSE is used to measure the average deviation between the output signal and the original true signal. It is the most direct indicator reflecting prediction error, reconstruction error, or denoising error. The smaller the RMSE value, the closer the model output is to the original signal, and the smaller the overall error.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2} \tag{21}$$

1) When the RMSE is small, it means that the model can effectively preserve the overall shape and trend of the original signal and will not produce obvious distortion during the reconstruction process.

2) When the RMSE is large, it indicates that the model error is significant, which may lead to signal reconstruction distortion, poor filtering effect or inaccurate prediction.

2. PRD measures the ratio of the error signal energy to the original signal energy, and is a relative error metric. Compared to RMSE, which focuses solely on error magnitude and may be affected by signal amplitude, PRD can provide a fairer performance comparison of signals with different amplitudes or energies through normalization.

$$PRD = \frac{\sqrt{\sum_{i=1}^{N}\left(x_i - \hat{x}_i\right)^2}}{\sqrt{\sum_{i=1}^{N} x_i^2}} \times 100 \tag{22}$$

1) When the PRD is low, it indicates that the proportion of error energy is small, and the model output is almost identical to the real signal, with high fidelity.

2) When PRD is relatively high, it indicates that the model has lost a significant amount of useful information during reconstruction or denoising, resulting in poor recovery quality.

3. SNR measures the ratio of the energy of useful information in the reconstructed signal to the energy of the error signal, and is a key indicator for evaluating a model's noise suppression capability and signal fidelity. SNR reflects whether a model can reduce noise interference while retaining the details of the useful signal.

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N}\left(x_i - \hat{x}_i\right)^2}\right) \tag{23}$$

1) The higher the SNR, the lower the error energy, the higher the quality of the reconstructed signal, and the more effectively the model can extract the true signal content.

2) A lower SNR indicates a higher noise component, meaning the model has not effectively suppressed the noise.

Considering the three metrics RMSE, PRD, and SNR, when RMSE and PRD significantly decrease, it indicates that the model effectively reduces errors and enhances signal fidelity; when SNR increases, it indicates that the model can better distinguish between valid signals and noise, thereby achieving higher quality signal reconstruction and denoising.

4.2 Analysis of Experimental Results

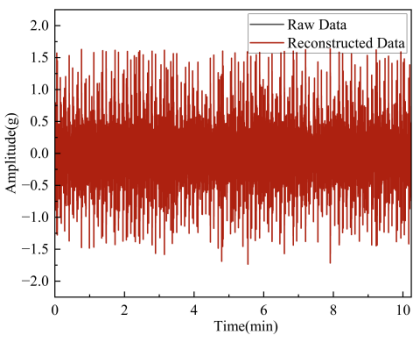


Figure a H_20Hz

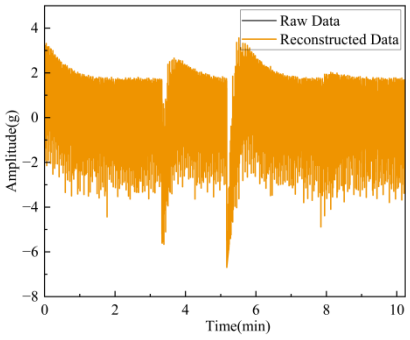


Figure b H_40Hz

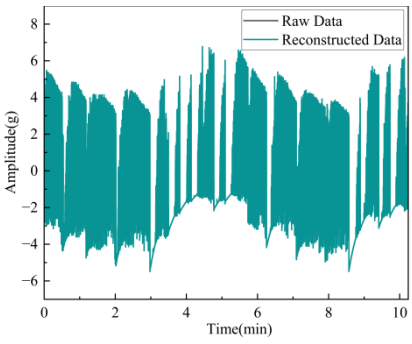


Figure c H_60Hz

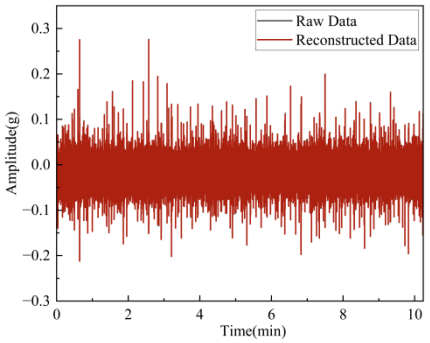


Figure a B_20Hz

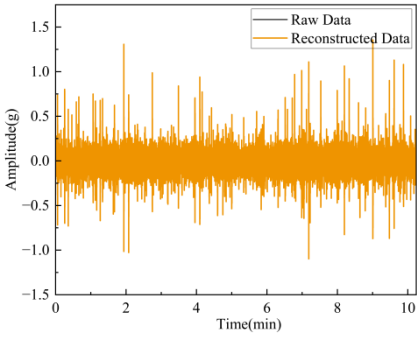


Figure b B_40Hz

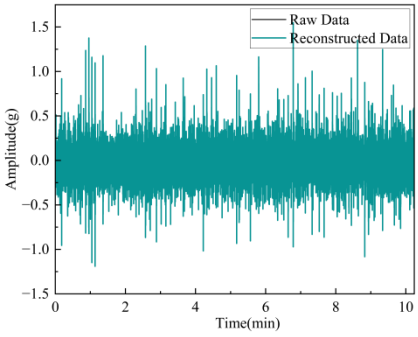


Figure c B_60Hz

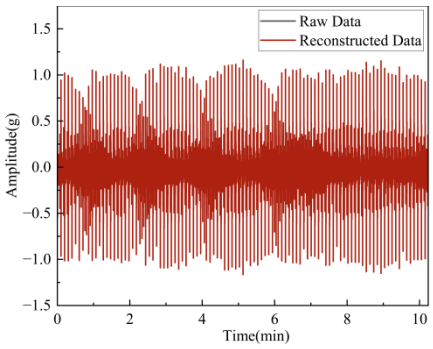


Figure a I_20Hz

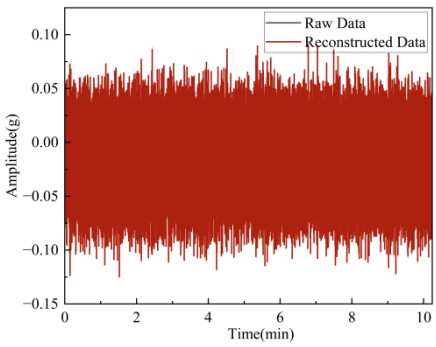


Figure a O_20Hz

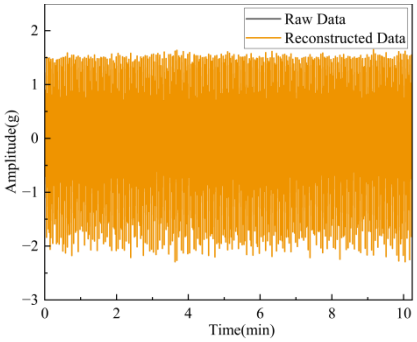


Figure b I_40Hz

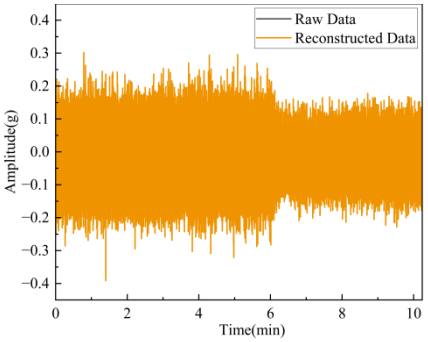


Figure b O_40Hz

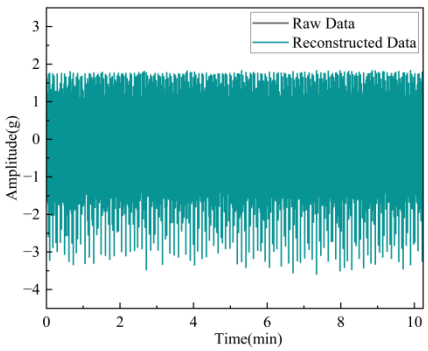


Figure c I_60Hz

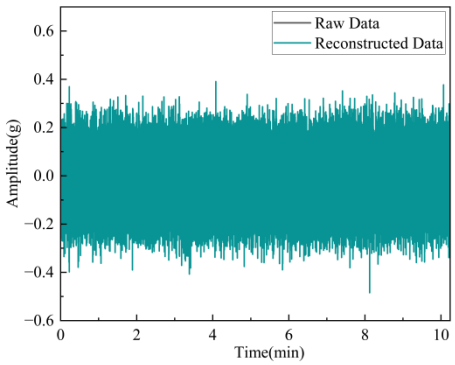


Figure c O_60Hz

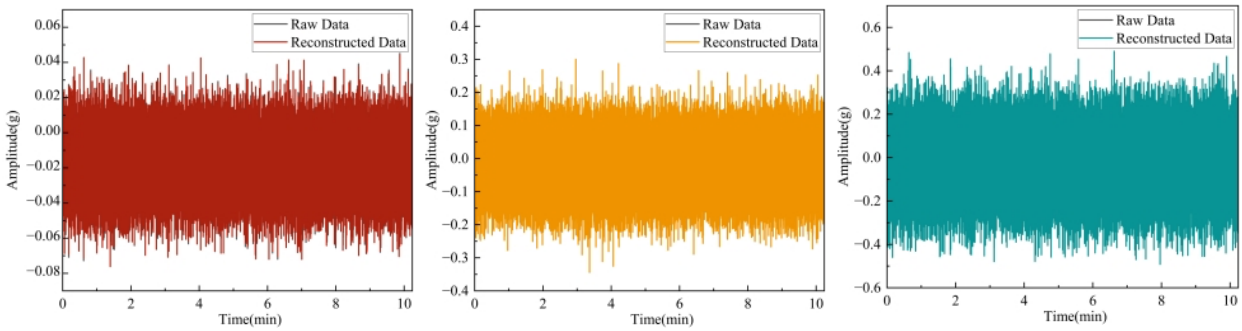


Figure a C_20Hz Figure b C_40Hz Figure c C_60Hz

To verify the HAGAN framework's ability to reconstruct vibration signals under different bearing conditions, a comparative analysis was conducted using normal conditions and various typical bearing fault conditions, including normal operation (H), rolling element fault (B), inner race fault (I), outer race fault (O), and combined fault (C). The time-domain characteristics of each type of signal and their reconstruction performance are analyzed as follows.

1.H: Normal operating conditions

As shown in the figure for the H-type operating condition, this state represents the motor bearing operating under normal conditions. The vibration signal in this state has a relatively small overall amplitude, the fluctuation process is relatively stable, and no significant impact components caused by faults are present. The vibration is mainly dominated by the basic noise generated during the equipment's operation. It can be observed that under the conditions of 20 Hz, 40 Hz, and 60 Hz, the overall waveform of the original signal is relatively stable, with no obvious impact components. The amplitude mainly exhibits small random fluctuations and only shows a slight increase as the frequency rises. The signals reconstructed by HAGAN closely match the overall waveform of the original signals at all frequencies, without introducing additional shocks or abnormal fluctuations. They effectively preserve the stability and continuity characteristics of normal operating signals, indicating that this method has good signal retention capability under fault-free conditions.

2.B: Rolling element (ball) fault

In the figure, Class B operating conditions correspond to faults in the motor bearing rolling elements (balls), which are usually caused by surface wear or local dents on the rolling elements. During the operation of the bearing, its vibration signal is mainly composed of random impact components, with the impacts being quite dispersed, making it difficult to form distinct periodic concentrated pulse characteristics. At the same time, as the operating frequency increases, the range of signal amplitude fluctuations shows a gradually expanding trend. As can be seen from the figure, this type of signal exhibits obvious random impact characteristics at all three frequencies, with a relatively dispersed impact distribution and a lack of stable periodic pulse structure; as the operating frequency increases from 20 Hz to 60 Hz, the amplitude fluctuation range expands significantly, and the signal's randomness is further enhanced. By comparing the original signal with the reconstructed signal, it can be observed that the HAGAN method effectively preserves the locations and amplitude variations of random impacts under different frequency conditions. The reconstructed waveform closely matches the original signal in both overall fluctuation trends and local impact details, indicating that this method has a strong capability to reconstruct the non-stationary and random impact features in rolling element faults.

3.I: Inner ring fault

Type I operating condition represents a fault in the motor bearing inner ring, meaning there are structural defects such as wear, local spalling, or dents on the bearing inner raceway. Since the inner ring rotates synchronously with the shaft, its defects periodically come into contact with the rolling elements during operation, thereby generating pulse impact features with a stable period in the vibration signal. This feature is an important basis for identifying inner ring faults. It can be clearly observed from the figure that this type of signal exhibits relatively stable periodic impact characteristics under all frequency conditions, with fairly regular impact intervals; as the frequency increases, both the impact density and amplitude increase. After

processing with the HAGAN method, the reconstructed signal remains highly consistent with the original signal in terms of the location of impacts, periodic structure, and amplitude variations. The periodic pulse characteristics are clearly distinguishable, with no obvious misalignment or attenuation, indicating that this method can effectively preserve the strongly regular periodic impact features in inner ring fault signals, with stable and reliable restoration results.

4.O: Outer ring fault

The O-type operating condition in the figure corresponds to an outer ring fault of the motor bearing. Because the outer ring is relatively stationary, its defects generate a position-stable impact response each time the rolling elements pass by, resulting in periodic pulse features in the original signal that are highly regular and repetitive. As the operating frequency increases, the impact amplitude increases overall. As can be seen from the comparison results, the HAGAN reconstructed signal can accurately restore the periodic impact structure of outer ring faults under various frequency conditions. The pulse positions are stable, the shapes are clear, and the reconstructed waveform highly coincides with the original signal in terms of time-domain structure, demonstrating the method's good ability to preserve the features of regular fault signals.

5.C: Combination Fault

Condition C represents a combined fault in the motor bearing, meaning that at least two types of bearing components, such as the inner ring, outer ring, and rolling elements, fail simultaneously. Under this condition, the vibration signal exhibits characteristics of multiple faults superimposed on each other. As can be seen from the figure, this type of signal contains both randomly dispersed impact components and relatively concentrated periodic pulse characteristics, representing a superposition of multiple single fault features; moreover, as the operating frequency increases, the range of amplitude fluctuations significantly expands, and the signal structure becomes more complex. The HAGAN reconstruction results retain the mixed impact characteristics in the combined fault signals relatively completely under different frequency conditions. The reconstructed curves are highly consistent with the original signals in terms of impact distribution, amplitude changes, and local fluctuation details, indicating that this method still has strong feature retention capability and stable recovery performance under complex operating conditions.

The results of these evaluation metrics across all 24 test scenarios are summarized in the following table, providing a detailed comparison of the model's performance.

Table 4: Performance metrics of the HAGAN model for signal reconstruction

Condition	Frequency (Hz)	RMSE ($\times 10^{-3}$)	PRD (%)	SNR (dB)
H	20	1.155	5.188	25.700
	40	1.154	1.589	35.976
	60	1.155	0.992	40.067
	80	1.155	0.241	52.371
I	20	1.155	1.575	36.054
	40	1.155	0.528	45.550
	60	1.155	0.301	50.437
	80	1.154	0.182	54.809
O	20	1.156	4.343	27.243
	40	1.155	1.919	34.340
	60	1.154	1.403	37.056
	80	1.155	0.328	49.680
B	20	1.154	4.657	26.638
	40	1.154	1.537	36.267
	60	1.154	1.220	38.270
	80	1.155	0.303	50.384
C	20	1.155	0.826	41.664
	40	1.154	0.157	56.057

	60	1.155	0.057	64.850
	80	1.155	0.054	65.380
M	20	1.155	5.256	25.586
	40	1.153	2.998	30.464
	60	1.155	1.970	34.112
	80	1.155	0.594	44.521

Note: H: Healthy; I: Inner race Fault; O: Outer race Fault; B: Ball Fault; C: Combination Fault; M: Moderate Fault.

Based on the provided evaluation metrics (RMSE, PRD, and SNR), the performance of the HAGAN model in reconstructing bearing vibration signals can be assessed across different fault conditions and frequencies. Taking the inner race fault (I) as an example, the RMSE values remain nearly constant at approximately 1.154×10^{-3} across all tested frequencies, indicating stable reconstruction accuracy. Meanwhile, as the frequency increases from 20 Hz to 80 Hz, the PRD decreases from 1.575% to 0.182%, and the corresponding SNR improves from 36.054 dB to 54.809 dB. These results demonstrate that the HAGAN model achieves more accurate signal reconstruction and better noise suppression under higher-frequency conditions for inner race fault signals.

5. Conclusion

This paper focuses on the problem that traditional compressed sensing and deep learning–based compression methods suffer from significant reconstruction performance degradation and loss of critical fault features when applied to motor bearing vibration signals under extremely high compression ratios. In particular, existing approaches often face a trade-off between compression efficiency, reconstruction accuracy, and computational complexity, which limits their applicability in real-time industrial fault diagnosis scenarios. To address these challenges, a lightweight hybrid compressed sensing framework based on an autoencoder and generative adversarial network, referred to as HAGAN, is proposed.

The proposed method integrates an autoencoder to achieve high-dimensional signal compression and a generative adversarial network to enhance reconstruction fidelity through adversarial learning. By simplifying the network structure and removing redundant nonlinear components, the framework effectively reduces computational overhead while preserving essential fault-related features. This design enables efficient signal compression and high-quality reconstruction, even under extreme measurement rate conditions.

To quantitatively evaluate the reconstruction performance of the proposed framework, three widely used evaluation metrics, namely RMSE, PRD, and SNR, are employed. RMSE is used to measure the overall reconstruction error, PRD reflects the relative energy of the reconstruction error, and SNR evaluates the signal fidelity and noise suppression capability of the reconstructed signal. Experimental results demonstrate that, across different bearing conditions and operating frequencies, the RMSE values remain stable at approximately 1.15×10^{-3}, indicating consistent reconstruction accuracy. Meanwhile, the PRD values decrease significantly with increasing operating frequency, reaching values below 1% in most cases, which implies that only a small proportion of signal energy is lost during reconstruction. Correspondingly, the SNR values show a clear increasing trend, with peak values exceeding 65 dB, demonstrating the effectiveness of the proposed method in suppressing noise while retaining useful signal components.

Under an extreme compression ratio of 100:1, the HAGAN framework is still able to accurately reconstruct vibration signals corresponding to normal operation and various fault conditions, including inner race faults, outer race faults, rolling element faults, and combined faults. The reconstructed signals preserve both periodic impact characteristics and randomly distributed transient components, indicating that the proposed method can maintain critical diagnostic features under severe data reduction. The joint analysis of RMSE, PRD, and SNR confirms that the proposed framework achieves a favorable balance between compression efficiency and reconstruction quality, outperforming traditional compressed sensing approaches in high-compression scenarios.

Although the proposed method achieves promising results in reconstructing bearing vibration signals under high compression ratios, the current experiments are primarily conducted on offline datasets. Further validation

is required to assess the real-time performance and long-term stability of the model in practical online monitoring scenarios.

References

[1] Moghadasian, M., Shakouhi, S. M. and Moosavi, S. S. Induction motor fault diagnosis using ANFIS based on vibration signal spectrum analysis. In 2017 3rd International Conference on Frontiers of Signal Processing (ICFSP), Paris, France, 2017; pp. 105-108. https://doi.org/10.1109/ICFSP.2017.8097151.

[2] Pu, H., Zhang, K. and An, Y. Restricted Sparse Networks for Rolling Bearing Fault Diagnosis. IEEE Transactions on Industrial Informatics. 2023, 19(11), pp. 11139-11149. https://doi.org/10.1109/TII.2023.3243929.

[3] Chen, X., Yang, R., Xue, Y., Huang, M., Ferrero, R. and Wang, Z. Deep Transfer Learning for Bearing Fault Diagnosis: A Systematic Review Since 2016. IEEE Transactions on Instrumentation and Measurement. 2023, 72, pp. 1-21. https://doi.org/10.1109/TIM.2023.3244237.

[4] Arie, R., Brand, A. and Engelberg, S. Compressive sensing and sub-Nyquist sampling. IEEE Instrumentation & Measurement Magazine. 2020, 23(2), pp. 94-101. https://doi.org/10.1109/MIM.2020.9062696.

[5] Jeong, J. Y., Ozger, M. and Lee, W. H. Compressed sensing vs. Auto-encoder: On the perspective of signal compression and restoration. IEEE Access. 2024, 12, pp. 41967-41979. https://doi.org/10.1109/ACCESS.2024.3366899.

[6] Ahmed, I., Khalil, A., Ahmed, I. and Frnda, J. Sparse Signal Representation, Sampling, and Recovery in Compressive Sensing Frameworks. IEEE Access. 2022, 10, pp. 85002-85018. https://doi.org/10.1109/ACCESS.2022.3197594.

[7] Ahmed, I., Wuttisittikulkij, L., Khan, A. and Iqbal, A. The Optimally Designed Deep Autoencoder-Based Compressive Sensing Framework for 1D and 2D Signals. IEEE Access. 2024, 12, pp. 150520-150539. https://doi.org/10.1109/ACCESS.2024.3472044.

[8] Pan, B., Shi, Z. and Xu, X. R-VCANet: A New Deep-Learning-Based Hyperspectral Image Classification Method. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2017, 10(5), pp. 1975-1986. https://doi.org/10.1109/JSTARS.2017.2655516.

[9] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Processing Magazine. 2012, 29(6), pp. 82-97. https://doi.org/10.1109/MSP.2012.2205597.

[10] Wu, H., Zheng, Z., Li, Y., Dai, W. and Xiong, H. Compressed Sensing via a Deep Convolutional Auto-encoder. In 2018 IEEE Visual Communications and Image Processing (VCIP), Taichung, Taiwan, 2018; pp. 1-4. https://doi.org/10.1109/VCIP.2018.8698640.

[11] Kulkarni, K., Lohit, S., Turaga, P., Kerviche, R. and Ashok, A. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In Proceedings of the IEEE conference on computer vision and pattern recognition, Vegas, NV, USA, 2016; pp. 449-458.

[12] Mousavi, A. and Baraniuk, R. G. Learning to invert: Signal recovery via deep convolutional networks. In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), New Orleans, LA, USA, 2017; pp. 2272-2276. https://doi.org/10.1109/ICASSP.2017.7952561.

[13] Zhang, J. and Ghanem, B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 2018; pp. 1828-1837.

[14] Chen, X., Chen, Y., Xie, Y. and Zhao, L. Learning Both Sensing Matrix and Sparse Representation via Deep Neural Network for Compressive Sensing. IEEE Transactions on Signal Processing. 2019, 67(23), pp. 6116–6129. https://doi.org/10.1109/TSP.2019.2946982.

[15] Puttegowda, K., Veeraprathap, V., Sequeira, S. S., Aruna, B., Sunil Kumar, D. S. and Sudheesh, K. V. Optimized Deep Learning Approach for Image Compression Using Compressed Sensing Technique with Convolutional Autoencoders. In 2024 International Conference on Data Science and Network Security (ICDSNS), Tiptur, India, 2024; pp. 1-5. https://doi.org/10.1109/ICDSNS62112.2024.10691037.

[16] Chen, P., Song, H., Zeng, Y., Guo, X. and Tang, C. A real-time and robust neural network model for low-measurement-rate compressed-sensing image reconstruction. Entropy. 2023, 25(12), p. 1648. https://doi.org/10.3390/e25121648.

[17] Zhao, C., Zio, E. and Shen, W. Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study. Reliability Engineering & System Safety. 2024, 245, p. 109964. https://doi.org/10.1016/j.ress.2024.109964.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

Acknowledgment
This paper is an output of the science project.