Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2, 2132-2146
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i2.5050

Bearing fault diagnosis in high noise environment using multi-scale processing, channel-attention and feature-enhanced convolutional neural network model

Xiuyu Li[1,2], Shirley J. Tanjong[1*]
[1]Faculty of Engineering, Universiti Malaysia Sarawak, 94300, Kota Samarahan, Sarawak, Malaysia; jtshirley@unimas.my (S.J.T.).
[2]Intelligent Manufacturing and Automobile Academy, Anhui Business and Technology College, HeFei, 231131, China.

Abstract: This paper presents a model using deep learning techniques which includes Multi-scale processing, Channel attention, Feature enhancement, and anomaly Classification layers, referred to as MCFCNN, for bearing fault diagnosis in noisy industrial environments. The MCFCNN network combines multi-channel parallel convolution, effectively capturing spatial information, and introduces channel attention mechanisms to adaptively recalibrate channel-level feature responses. Secondary neurons are introduced to enhance the model's ability to capture complex nonlinear patterns related to bearing faults. The model was tested and compared to other models using a publicly available data set. In a simulated high-noise environment, the proposed model outperforms existing models in fault diagnosis, with accuracy greater than 80% even at high signal-to-noise (SNR) ratio. At SNR = -6, the MCFCNN records higher accuracy (83%), precision (89%), and recall rates (84.5%) as compared to prior models. The proposed model can be integrated into the maintenance management system to enhance bearing health assessment and prediction, improving machine prognostics.

Keywords: *Bearing fault diagnosis, Channel attention, Multi-scale processing, Quadratic convolutional neural.*

1. Introduction

Rotating machines are core parts of equipment in power plants and industrial production. Bearings are key components that support and maintain rotating parts and are vital in ensuring the smooth operation and long-term reliability of machinery. However, owing to the complex working environment, high loads, and continuous operation of the rotating machinery, the failure rate is high. Faults in components of mechanical equipment may cause the entire system inoperable and failures of critical components can lead to significant casualties and economic losses (Rai & Upadhyay, 2016). Bearing failures account for a significant proportion of mechanical equipment failures (Qiu et al., 2023). Bearing fault signals can be collected by installing various types of sensors to monitor the working conditions of bearings, including vibration (H.-Y. Chen & Lee, 2020; Z. Chen, Wang, Zhou, Yang, & Liu, 2024) temperature (Eang & Lee, 2024; Gunerkar, Jalan, & Belgamwar, 2019) rotational speed (Eang & Lee, 2024) and electrical current (Bouras, Bennedjai, & Bouras, 2020) among others.

Vibration-based signal acquisition has been extensively researched and applied owing to its independence from the mechanical structure and simplicity of testing. However, the complexity of the environment and variability of the load result in the interference of noise in the collected vibration signals during the fault diagnosis process, increasing the difficulty of fault diagnosis. Figure 1 illustrates a vibration signal without noise and with noise.

History: Received: 20 December 2024; Revised: 23 January 2025; Accepted: 25 January 2025; Published: 26 February 2025
* Correspondence: jtshirley@unimas.my

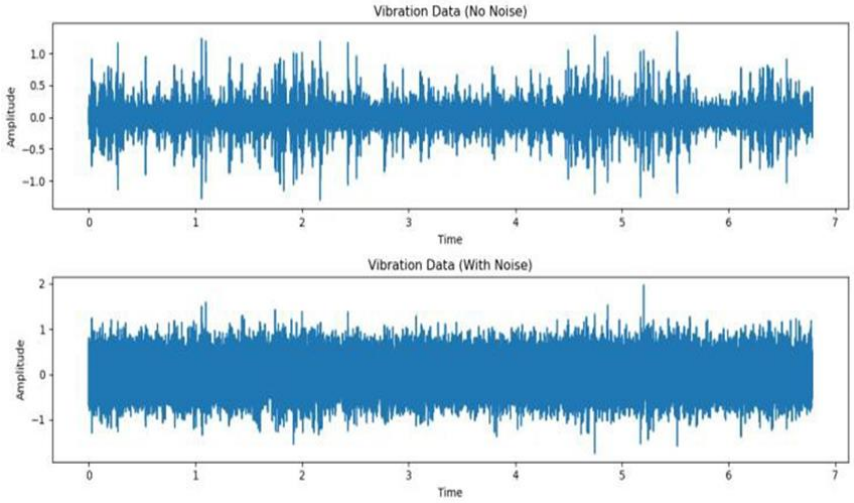

Figure 1.
Sample of vibration signals: (a) Top graph – vibration data without noise; (b) Bottom graph – vibration data with noise.

For bearing fault diagnosis in a noisy environment, traditional methods struggle to extract fault characteristic frequencies directly (Cerrada et al., 2018). Frequently, fault feature extraction is achieved through time-domain (Bousseksou, Bessous, & Mahmoud, 2024; Zhao, Zhu, & He, 2024) frequency-domain, and time–frequency signal processing methods, followed by inputting the extracted features into a diagnostic model for fault classification. Common feature-extraction methods include the short-time Fourier transform (T. Wang, Liang, Li, & Cheng, 2014) fast Fourier transform (G. Chen, Liu, & Chen, 2020) Hilbert transform (Dai et al., 2020) and wavelet transform (Duong & Kim, 2018). Common machine-learning diagnostic models include logistic regression (Shang et al., 2017) k-nearest neighbours (Lu et al., 2021) and support vector machines (M. Wang et al., 2021). However, traditional methods have some limitations: 1) They rely on manual feature extraction, which results in complex signal processing designs and a heavy reliance on expert experience. When the operating conditions are complex, these methods fail to address the problem of differences in data distribution, posing significant challenges for fault diagnosis. 2) The diagnostic model requires laborious design steps and different model structure parameters are required for different fault types. When signals are contaminated by noise, the diagnostic performance is suboptimal.

Therefore, current research on bearing fault diagnosis in noisy environments focuses on deep learning methods. Deep learning has deep structures and effective nonlinear feature-extraction capabilities, enabling it to extract fault features directly from raw data and achieve end-to-end fault diagnosis, avoiding the tedious feature-extraction and classification steps of traditional methods. Deep learning methods can handle large-scale datasets and complex signal conditions and exhibit robustness and generalisation performance. The vibration signals collected for bearing diagnostics are generally one-dimensional time series signals. Currently, two methods are used for bearing diagnostics based on the input dimensions of the network.

One approach is to directly process one-dimensional data. The 1DCNN model proposed in Eren, Ince, and Kiranyaz (2019); Ozcan, Devecioglu, Ince, Eren, and Askar (2022) and Huan Wang, Liu, Peng, and Qin (2019) has been proven to have highly generalizable performance for bearing fault diagnosis. Another approach involves the transformation of one-dimensional data into two-dimensional data for further processing. Some studies converted the time-domain vibration signals of bearings into grayscale images containing fault features, which were then input into a 2DCNN model (Wan, Chen, Li, & Li, 2020; Y. Wang, Ding, Zeng, Wang, & Shao, 2020; Xu, Liu, Jiang, Söffker, & Shen, 2019). The trained

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

model can effectively recognise the types of faults and their severity. In comparison, the 1DCNN network that directly processes one-dimensional data is simpler, whereas converting from one-dimensional to two-dimensional data may result in some information loss.

Therefore, a W1DCNN model was proposed in Zhang, Peng, Li, Chen, and Zhang (2017) and it has two main characteristics. The first layer uses a wide 64-channel convolutional kernel, whereas the other layers use a network structure with small 3×3 convolutional kernels. A high accuracy was achieved with this model. Based on the success of the W1DCNN in diagnostics, researchers have proposed several improved versions of it Y. Wang et al. (2020) and Shenfield and Howarth (2020). Recently, owing to the effective feature-representation capabilities of quadratic neural networks (Fan, Xiong, & Wang, 2020; J.-X. Liao et al., 2023), the traditional linear neural network convolution in the W1DCNN network has been replaced with quadratic neural elements, resulting in a QCNN model. Many researchers have introduced attention mechanisms into bearing fault diagnosis tasks, enabling traditional CNN models to achieve higher classification accuracy (Huan Wang et al., 2019; Hui Wang, Xu, Yan, Sun, & Chen, 2020).

Although deep learning has achieved significant success in bearing fault diagnosis research, current deep learning models have some limitations. Many deep learning models perform well on clean signals with consistent data distributions. However, when the signals contain a large amount of noise, the classification performance of the models decreases significantly. For example, when a signal-to-noise ratio (SNR) of -6 is added to the CWRU dataset, the classification accuracies of AResNet (Zhong, Wang, & Ban, 2022) W1DCNN (Zhang et al., 2017) and QCNN (J.-X. Liao et al., 2023) are 61.2%, 66.5%, and 71.1%, respectively. These models require further improvements in their classification accuracy in noisy environments. Noise is ubiquitous in industrial settings, making it crucial to efficiently diagnose signals that contain multiple sources of noise.

Based on the aforementioned problems, this paper proposes a new network architecture called MCFCNN. The network model contains multi-scale processing, channel attention, feature enhancement, and anomaly classification layers, which are used for the efficient diagnosis of bearing faults in multi-noise environments.

2. Methods

This paper utilises time-series vibration data collected during the operation of bearings for fault diagnosis. The diagnostic problem can be defined as follows: the input is $X_t \in R^k$, where $t = (0, 1, 2, \ldots, T)$. Here, T is the time step, and k is the data type, which includes both fault data and healthy data of the bearings. The corresponding output is the fault type Y_t, for model fault diagnosis. The goal is to achieve fault diagnosis by establishing a mapping relationship between Y_t and the dataset, as follows:

$$Y_t = f(X) \tag{1}$$

where X is the bearing operation data, f is the mapping function, and Y_t is the fault classification result of f diagnosis.

The proposed bearing fault diagnosis network is shown in Figure 2. The entire network consists of four layers: multi-scale processing, channel attention, feature enhancement, and anomaly classification layers.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

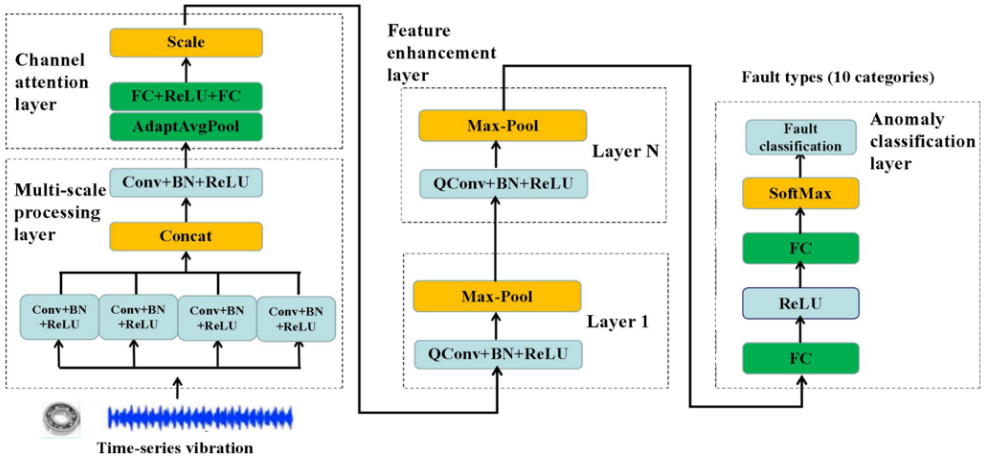

Figure 2.
MCFCNN network architecture.

2.1. Multi-scale Processing Layer

The multi-scale processing layer comprises a parallel convolution operation that utilises multiple parallel convolution branches to extract different features. Each branch can use different convolution kernels to process input data and generate a set of feature maps. These feature maps extract features from different aspects and are then fused to obtain a more comprehensive feature representation.

Assuming the input data are $\boldsymbol{X}$, the number of convolution kernels is $\boldsymbol{N}$, the size of the convolution kernel is $\boldsymbol{K \times K}$, the number of channels in the input data is $\boldsymbol{C}$, and the number of channels in the output feature maps is $\boldsymbol{M}$. For each convolution kernel $\boldsymbol{i}$ $(\boldsymbol{1 \leq i \leq N})$, its weight matrix is $\boldsymbol{W_i}$ with a shape of $\boldsymbol{K \times K \times C}$, and the bias term is $\boldsymbol{b_i}$. The output of the multi-branch convolution operation for the m-th feature map of the input data $\boldsymbol{X}$ can be calculated using the following formula:

$$Y_m = \sum^n f(conv(W_n, X) + b_n) \qquad (2)$$

where $\boldsymbol{conv(W_n, X)}$ represents the convolution operation between the input data $\boldsymbol{X}$ and the convolutional kernel $\boldsymbol{W_n}$; $\boldsymbol{b_n}$ represents the bias term of the $\boldsymbol{n}$-th convolutional kernel; and $\boldsymbol{f}$ represents the rectified linear activation function (ReLU) function.

In bearing fault diagnosis, different types of faults exhibit different fault characteristic frequencies, which result in different sensitivities of different scales of features to different types of faults. Through multi-branch convolutional operations, the vibration features at different scales can be captured simultaneously, thereby improving the detection and diagnostic capabilities of the model for bearing faults. This enhances the adaptability and generalisation ability of the model for bearing faults. Additionally, feature fusion at different scales can acquire a more comprehensive and rich feature representation. This aids the model in better understanding the vibration signals at different scales and improves its discriminative power for bearing faults.

2.2. Channel Attention Layer

The channel attention network by Bahdanau, Cho, and Bengio (2014) won the championship in the ImageNet competition for classification tasks in 2017. It can be used to enhance the attention mechanism of convolutional neural networks. It improves network performance by adaptively adjusting the importance of features across channels.

The channel attention module consists of two key steps i.e. squeezing and excitation (Figure 3).

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

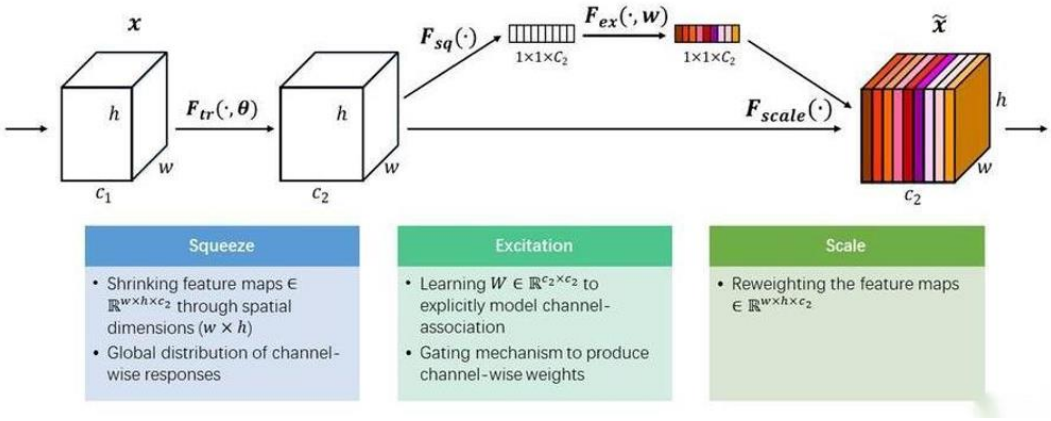

Figure 3.
Channel Attention Module
Source: Hu, Shen, and Sun (2018).

Squeezing involves global pooling on the input feature map along the channel dimensions to obtain global information regarding the channels. Assuming that the size of the input feature map is $(\boldsymbol{H},\boldsymbol{W},\boldsymbol{C})$, the squeeze operation performs a global average pooling on the input feature map, resulting in a $\boldsymbol{C}$-dimensional vector that represents the global average value for each channel. This vector represents the overall importance of each channel in the input feature map. The resulting output Z can be expressed as

$$Z = \mathrm{F}_{sq}(\mathrm{X}) = \frac{1}{H\times W}\sum_{i=1}^{H}\ \sum_{j=1}^{W}\ X(i,j) \tag{3}$$

Excitation then uses a learned nonlinear function to transform the squeezed features into channel attention weights, which are used to adjust the importance of each channel in the input feature map. The excitation operation generates the channel attention weights by learning a nonlinear mapping function. This mapping function can be learned using either fully connected or one-dimensional convolutional layers. Given an input feature vector $\boldsymbol{Z}$, the result of the excitation operation $\boldsymbol{S}$ can be expressed as

$$S\ =\ \sigma(W2\ *\ \delta(W1\ *\ Z)) \tag{4}$$

S is a C-dimensional vector representing the attention weights for each channel; W1 and W2 are weight matrices; δ represents the ReLU activation function; and σ is the sigmoid function.

Scale: By performing an element-wise multiplication between S and X, the channels of X can be reweighted. Therefore, the value of the weighted feature map Y is

$$Y\ =\ F_{scale}(S,X)\ =\ S*X \tag{5}$$

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

The channel attention layer can adaptively learn the correlations between feature channels and enhance the important features in bearing vibration signals. In bearing fault diagnosis, samples of faulty cases are often limited, which means that the model must learn from limited data and generalise to unseen fault scenarios. The channel attention module can enhance the sensitivity and discriminability of the model for bearing fault samples. By adaptively adjusting the weights of the feature channels, the model can focus more on features relevant to bearing faults, enabling a more accurate identification and classification of faulty samples.

2.3. Feature Enhancement Layer

Traditional neural network models commonly use activation functions such as ReLU and sigmoid to capture non-linear relationships. However, these functions can only capture first-order features. Fan et al. (2020) proposed a novel neural network that used quadratic neurones to enhance the expressive power of a model and capture higher-order nonlinear features. This promotes neuronal diversity in deep learning. The distinctive feature of this approach is the integration of the two inner products of the input vector and the power term before the activation function. If the input vector is x, the output of the quadratic neuron, denoted as $\boldsymbol{f(x)}$, is

$$\sigma\big(f(x)\big) = \sigma\left(\left(\sum_{i=1}^{n} w_i^r x_i + b^r\right)\left(\sum_{i=1}^{n} w_i^g x_i + b^g\right) + \sum_{i=1}^{n} w_i^b x_i^2 + c\right)$$

$$= \boldsymbol{\sigma}((\boldsymbol{x}^\top \boldsymbol{w^r} + \boldsymbol{b^r})(\boldsymbol{x}^\top \boldsymbol{w^g} + \boldsymbol{b^g}) + \big(\boldsymbol{x} \odot \boldsymbol{x}\big)^\top \boldsymbol{w^b} + \boldsymbol{c}\big) \quad (6)$$

σ represents the activation function; ⊙ denotes the Hadamard product; $\boldsymbol{w_r}$, $\boldsymbol{w_g}$, and $\boldsymbol{w_b}$ are weight matrices; $\boldsymbol{b_r}$, $\boldsymbol{b_g}$, and $\boldsymbol{c}$ are bias vectors.

Quadratic neural networks have higher feature-extraction capabilities Fan et al. (2020). J.-X. Liao et al. (2023) explained that the mathematical formulation of quadratic neurones can derive a module similar to an attention mechanism, which aids the network in focusing more on local information. Bearing fault signals, which are generally high frequency, short range, and high amplitude, belong to local feature information. Therefore, in bearing fault diagnosis tasks, particularly in complex noisy environments, we choose to replace traditional linear convolution operations with quadratic neural network convolutions.

2.4. Anomaly Classification Layer

After a series of feature extractions, the vibration signal of the bearing is mapped to the sample label space using an anomaly classification layer to achieve fault classification. The MCFCNN model employs two linear layers to reduce the dimensionality of the features to 10 and then performs fault feature classification using the $\boldsymbol{SoftMax}$ function. Specifically, assuming that the input vector is $\boldsymbol{x}$, the linear layer $\boldsymbol{h(x)}$ can be expressed as

$$h(x) = f(wx + b) \quad (7)$$

The output of the anomaly classification layer $Y(x)$ of the MCFCNN can be expressed as

$$Y(x) = SoftMax(f(h(x)w_1 + b_1)) \quad (8)$$

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

where $\boldsymbol{w}$ and $\boldsymbol{w_1}$ are weight matrices, $\boldsymbol{b}$ and $\boldsymbol{b_1}$ in equations (7) and (8) are bias vectors; and h(x) represents the output of the first layer.

3. Experiments

A publicly available bearing dataset is used to validate the performance of the MCFCNN model. The proposed method is compared with other methods under different noise conditions.

3.1. Dataset Description

The Case Western Reserve University Bearing Dataset (CWRU) (Achen/Case, 2024) is a publicly available dataset for bearing fault diagnosis, created in 2003 by Dr. James T. Allison and colleagues from the Department of Mechanical Engineering at Case Western Reserve University in the United States. This dataset consists of accelerometer signals from four bearings, each containing data from normal operation as well as various types of fault conditions. The fault types include inner race faults, outer race faults, and rolling element faults. This dataset is utilised for research and experimentation in mechanical fault diagnosis and predictive analytics by many other researchers such as in Jalayer, Jalayer, Mor, Orsenigo, and Vercellis (2024); (Y. Liao et al., 2023); Rosa, Braga, and Silva (2024) and Yoo, Jo, and Ban (2023). It contains extensive time-series signals suitable for studying the performance of vibration analysis and signal processing algorithms.

The test rig consisted of a 2-horsepower motor, a torque sensor/encoder, and a dynamometer. The tested bearings supported the motor shaft. Faults of different sizes were induced on the inner race, rolling elements, and outer race using electrical discharge machining. The vibration data were collected at two sampling rates, 12 and 48 kHz, from both the fan and drive ends of the motor. Additionally, the data included four load conditions (0, 1, 2, and 3 HP), with motor speeds varying between 1730 and 1797 rpm, depending on the load. In this study, we utilized vibration signals collected at the fan end at a sampling rate of 12 kHz. The dataset consists of acceleration signals from four bearings, each containing data for the normal state and different types of fault states. The fault types include inner race, outer race, and rolling element faults. Each fault type has three diameters. Therefore, the dataset includes nine types of faulty bearings and one healthy bearing. The fault types and labels are summarized in Table 1.

Table 1.
Ten classes of bearing condition in CWRU data set.

Label	Fault Mode	Label	Fault Mode
0	Ball (Slight)	5	Inner Race (Serious)
1	Ball (Medium)	6	Outer Race (Slight)
2	Ball (Serious)	7	Outer Race (Medium)
3	Inner Race (Slight)	8	Outer Race (Serious)
4	Inner Race (Medium)	9	Healthy

3.2. Experimental Setup

CWRU dataset contains very long vibration signals, thus each sample was defined as a sequence with a length of 2048. Data augmentation techniques were applied to randomly crop 1000 samples to form a dataset for each sample. The total number of samples was 10,000. All samples were normalised. Finally, the dataset was split into training, validation, and testing sets at a ratio of 0.5:0.25:0.25.

The loss function used for training was cross-entropy, and the network was optimised using a stochastic gradient descent (SGD) optimiser. The gradient clipping technique was applied during the iteration to prevent a gradient explosion. The batch size for all the networks was set to 64, the learning rate was 0.1, and the momentum was 0.9. Through the iterative training of the model, the specific architecture of the MCFCNN network was defined, as shown in Table 2.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

Table 2.
The architecture configuration of MCFCNN.

Layer	Type	Kernel	Channel	Stride	Padding	Output
0	Input	-	-	-	-	1x2048
1	Conv+BN+ReLU(4)	3x1	4	1	1	4x2048
2	Conv+BN+ReLU	3x1	32	1	1	32x2048
3	Squeeze	-	-	-	-	32x1
4	Excitation	-	-	-	-	32x2048
5	QConv+BN+ReLU	64x1	16	8	28	16x256
6	MaxPool	2x1	-	2	-	16x128
7	QConv+BN+ReLU	3x1	32	1	1	32x128
8	MaxPool	2x1	-	2	-	32x64
9	QConv+BN+ReLU	3x1	64	1	1	64x64
10	MaxPool	2x1	-	2	-	64x32
11	QConv+BN+ReLU	3x1	64	1	1	64x32
12	MaxPool	2x1	-	2	-	64x16
13	QConv+BN+ReLU	3x1	64	1	1	64x16
14	MaxPool	2x1	-	2	-	64x8
15	QConv+BN+ReLU	3x1	64	1	-	64x6
16	MaxPool	2x1	-	2	-	64x3
17	Linear+ReLU	-	-	-	-	192x100
18	Linear	-	-	-	-	100x10
19	Output	10 classes				

The feature enhancement layer consisted of six layers: the first layer used a large convolutional kernel with a width of 64, and the remaining five layers used small convolutional kernels with a width of 3. This architecture was obtained from the W1DCNN (Zhang et al., 2017).

All experiments were conducted on a Windows 10 operating system equipped with an Intel i5-12490F CPU and an NVIDIA RTX 3060 12 GB GPU. Our code was written using Python 3.8 and PyTorch.

In this study, three performance metrics (accuracy, precision, and recall) were employed to evaluate the diagnostic performance for faults. Their definitions are as follows:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (9)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (10)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (11)$$

TP, FP, TN, and FN mean true positive, false positive, true negative, and false negative, respectively.

Gaussian white noise was added to the dataset and the performance metrics of the model under noisy conditions were evaluated. The SNR is defined as follows:

$$\mathrm{SNR_{dB}} = 10\log_{10}\left(\frac{\mathrm{P_{signal}}}{\mathrm{P_{noise}}}\right) \quad (12)$$

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

P_{signal} and P_{noise} are the powers of the signal and noise, respectively.

3.3. Comparison with Other Models

Accuracy: Performance of the MCFCNN model in different SNR environments is compared with the W1DCNN and QCNN models using the same data preprocessing techniques. Results presented in Table 3 compares the recognition accuracy under seven different SNR conditions, and all the results were averages of 10 runs. The best results among the three models are highlighted in bold, indicating that the MCFCNN model outperformed the other two models in almost every SNR environment.

Table 3.
Accuracy of different models at different SNRs.

	SNR	-6	-4	-2	0	2	4
CWRU (0HP)	W1DCNN	66.52%	77.36%	87.12%	94.20%	98.36%	99.52%
	QCNN	71.12%	79.92%	89.28%	95.76%	98.52%	99.64%
	MCFCNN	81.40%	91.04%	96.84%	99.72%	99.92%	99.96%
CWRU (1HP)	W1DCNN	69.68%	72.68%	88.60%	92.28%	98.12%	97.92%
	QCNN	70.80%	81.44%	89.82%	96.24%	98.72%	99.40%
	MCFCNN	81.88%	92.32%	97.44%	96.64%	98.84%	99.16%
CWRU (2HP)	W1DCNN	71.92%	83.72%	90.76%	97.80%	99.08%	99.84%
	QCNN	75.88%	85.16%	93.52%	97.96%	99.64%	100%
	MCFCNN	87.6%	97.28%	96.64%	99.72%	99.88%	100%
CWRU (3HP)	W1DCNN	72.40%	82.64%	90.68%	96.60%	99.68%	99.88%
	QCNN	73.96%	84.44%	92.68%	97.56%	99.52%	100%
	MCFCNN	87.66%	92.84%	98.12%	99.72%	100%	100%

Based on the diagnostic results in Table 3, we calculate the average accuracy of each model for the four operating conditions and plotted them as line graphs, as shown in Figure 4.

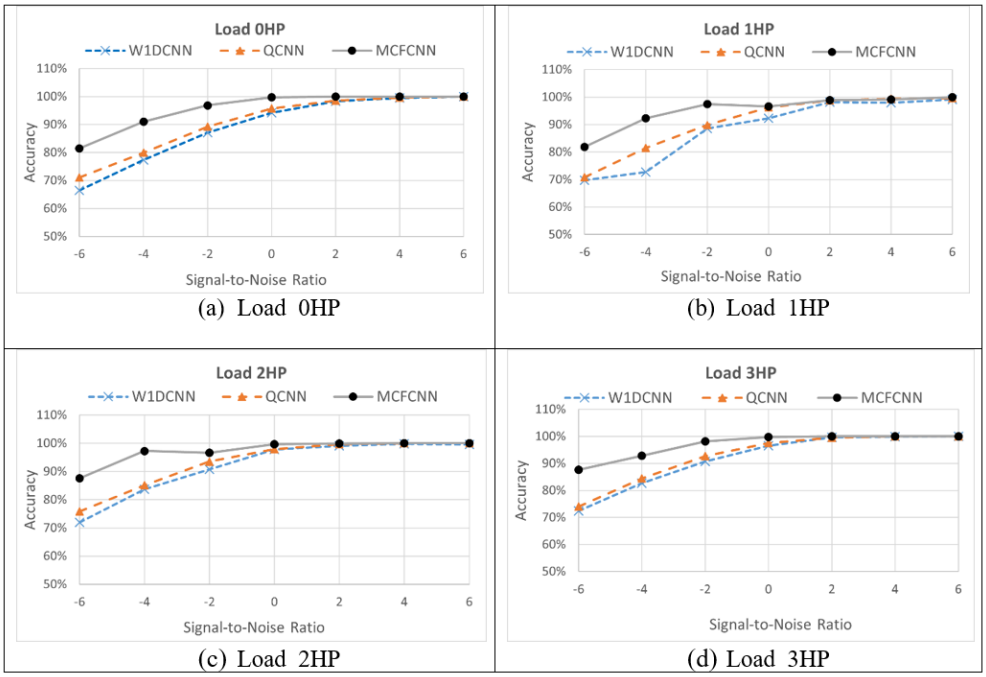

Figure 4.
Average accuracy of the model under different SNRs.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

Figure 4(a)-(d) describes that as the SNR increased from -6 to 6, the diagnostic accuracy of all three models improved. This implied that when the noise in the signal is relatively low, and the difference between the signal and background noise is more pronounced, the models are more likely to accurately identify and classify the faults. Figure 4(c) and (d) demonstrates that the MCFCNN model achieved an average accuracy 10-20% higher than the other two models when the SNR was -6. Additionally, when considering cases in which the diagnostic results were 100% accurate across all seven SNR levels, the W1DCNN achieved this only once, QCNN achieved it four times, and MCFCNN achieved it six times. Therefore, it can be concluded that the MCFCNN consistently provides high performance across various noise environments.

Figure 5 shows the training loss graphs for the three models under the 1 HP condition. The graph shows that the loss exhibited an overall decreasing trend. As the training progressed, the loss gradually decreased, indicating that the models were optimised during the learning process. Specifically, W1DCNN had the highest loss value, followed by QCNN, whereas MCFCNN had the lowest loss. This suggested that the MCFCNN model performed relatively well in terms of loss. The stability of the model can be preliminarily assessed by observing fluctuations in the loss. MCFCNN had relatively small fluctuations in loss, indicating that it is a more stable model with better adaptability to input data.

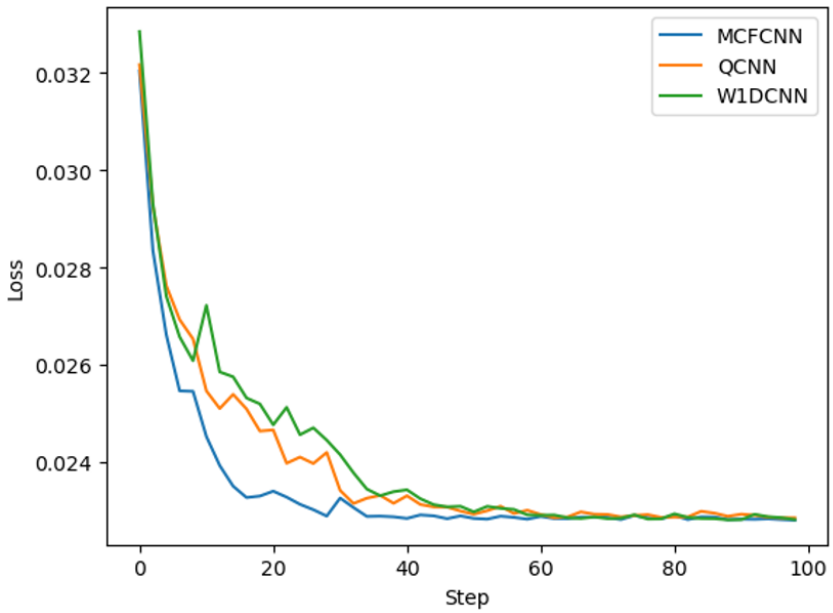

Figure 5.
Losses of different models (1 HP).

t-SNE Visualisation: When a signal is heavily disturbed by noise, the fault recognition of bearings is also compromised. For example, when $SNR = -6$ and the load was 0 HP, the classification accuracy of the W1DCNN model was 66.52% and that of QCNN was 71.12%; however, MCFCNN still maintained a high recognition rate of 81.40%. Because QCNN and MCFCNN have similar network architectures and quadratic neuron networks, the parallel convolution and SE modules in our proposed network MCFCNN, play a positive role in recognising fault features. To better compare the patterns learned by the three models in the training data, we output the results of the last convolutional layer and then reduced the high-dimensional data to 2D. The data as divided into ten fault categories and used t-SNE (Pouyet, Rohani, Katsaggelos, Cossairt, & Walton, 2018) for visualisation, as shown in Figure 6.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

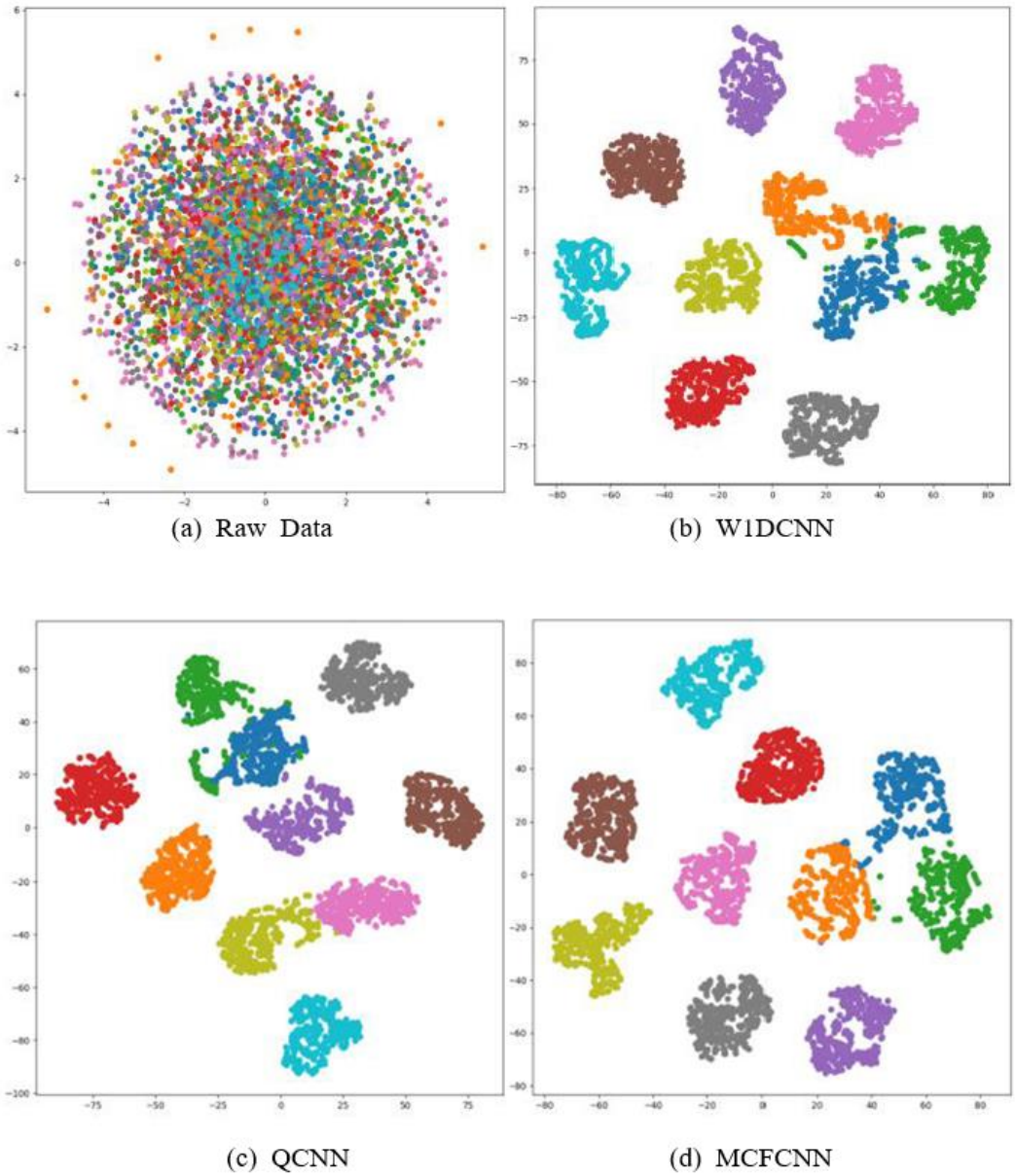

Figure 6.
Different models' t-SNE visualization results on the CWRU dataset (1HP, $SNR = -4$).

The visualisation results showed that all three models learned patterns of different categories, and the same fault categories were clustered. However, some overlaps occurred in the classification of certain categories. For example, in the W1DCNN results, the orange, blue, and green categories were not well distinguished at the boundaries, and the green category was scattered in clusters, indicating that this model did not learn well for the green category. In the QCNN results, some improvement was observed, as the orange category was completely separated, but the blue and green categories were not completely distinct, and the purple and dark yellow categories were close to each other at the boundaries. In the MCFCNN results, although some points in the blue and green categories were close

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

to the orange category, their numbers were relatively small, and overall, the boundaries of the three categories were distinguished.

Confusion matrix comparison: The classification performance of the fault modes under strong noise was analysed in detail by calculating the confusion matrices of the MCFCNN and QCNN models. The values on the diagonal of the confusion matrix represent the numbers of correctly identified instances. The numbers outside the diagonal represented the number of misclassified instances. Under the interference of strong noise, the QCNN model accurately identified four categories. The MCFCNN model identified five categories correctly. Both models accurately identified healthy bearings. Both models misclassified 52 instances as outer race faults, indicating errors in identifying the size of the outer race faults. For inner race faults, the MCFCNN model misclassified 20 instances, whereas the QCNN model misclassified 77 instances; the QCNN model mistakenly classified inner race faults as rolling element faults.

The performances of both models in classifying rolling element bearing faults were relatively low, primarily because of their inability to accurately identify the fault sizes of the rolling elements. The MCFCNN model (475/750 = 63.33%) accurately identified more rolling element faults than the QCNN model (333/750 = 44.4%). Finally, we were more interested in the number of misclassified samples for fault-type identification. Despite the strong noise interference, the MCFCNN model achieved 100% accuracy in recognising fault types, whereas the QCNN misidentified 77 samples of inner race faults as rolling element faults. All results indicated that the MCFCNN model performed better.

3.4. Analysis of MCFCNN

To determine which part of the model has a better ability to extract fault signals from the bearing, we decomposed the MCFCNN model. We separately added a multi-scale processing layer and channel attention layer to the network for comparison. For ease of identification, the network with only the multi-scale processing layer is referred to as MFCNN, and the network with only the channel attention layer is referred to as CFCNN.

According to Table 4, adding these two layers separately improved the classification performance of the model. Specifically, adding only the channel attention layer outperformed adding only the multi-scale processing layer by 3.46%. This indicated that in a noisy environment, the channel attention layer has a better capability of extracting fault features and assigning more weights to them. However, in a strongly noisy environment, the MCFCNN model exhibited higher accuracy, precision, and recall rates.

Table 4.
Comparison of model results (SNR = -6).

Methods	Accuracy	Precision	Recall
MFCNN	0.7716	0.7933	0.7908
CFCNN	0.8065	0.8215	0.8192
MCFCNN	0.8312	0.8899	0.8452

To investigate the patterns learned by the multi-scale processing and channel attention layers in network training, we defined them as the first layer of the MCFCNN network. Visualisation techniques were applied to observe the information focused on by the first layer of the network when fed noisy, healthy, and faulty bearing signals. Figure 6(a) shows the healthy bearing signal and Figure 6(b) the faulty bearing signal.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

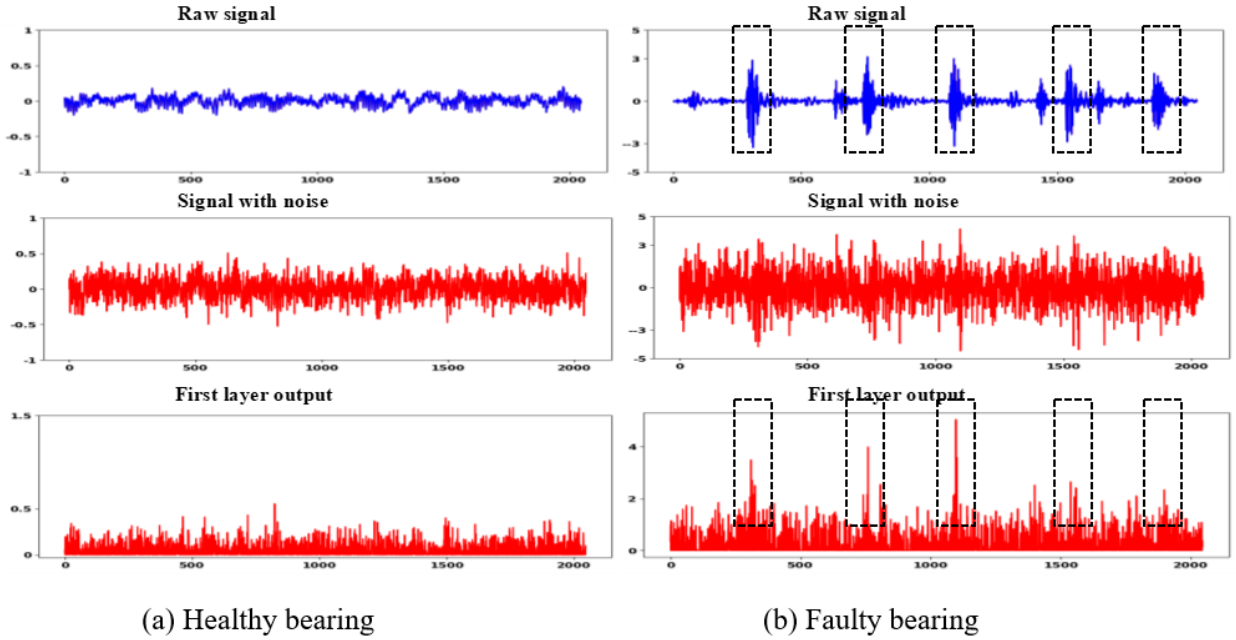

Figure 7.
Output feature map of the first layer of the MCFCNN.

Figure 7(a) shows the original signal, signal with added noise, and output signal after passing through the first layer of the MCFCNN. Figure 7(b) shows that despite the severe noise interference in the original signal, the network could still capture the location of the fault signal. This also explains why the MCFCNN maintained good classification performance even under noisy conditions.

4. Conclusions

In this article, we proposed a convolutional network called MCFCNN for end-to-end bearing fault diagnosis. The proposed model incorporates quadratic neuron convolutional network operations and introduces multi-scale processing and channel attention layers to enhance the diagnostic capability of the model in the presence of strong noise. The network achieved good classification performance on the CWRU-bearing dataset, maintaining an accuracy of over 80%, even in the presence of strong noise interference. Furthermore, classification tests were conducted on the multi-scale processing and channel attention layers to validate their importance in extracting bearing fault features. Visualisation techniques were employed to interpret the output feature maps. In the future, further improvements to the network will be required to enhance its practical applications.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgements:

This work was supported by the Natural Science Key Program of Anhui Province (2022AH052792), the Excellent Young Talents Support Program of Anhui Province University (gxyq2022274), and Universiti Malaysia Sarawak, Malaysia.

References
Achen/Case. (2024). *Achen/CaseWesternReserveUniversityData', Gitee*. Retrieved from https://gitee.com/zthcool/CaseWesternReserveUniversityData

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprintarXiv:1409.0473*.

Bouras, A., Bennedjai, S., & Bouras, S. (2020). Experimental detection of defects in variable speed fan bearing using stator current monitoring. *SN Applied Sciences*, *2*, 1-8. https://doi.org/10.1007/s42452-020-2687-2

Bousseksou, R., Bessous, N., & Mahmoud, M. M. (2024). Diagnosis of bearing faults using an advanced feature extraction based on WPT and EEV methods. *Studies in Engineering and Exact Sciences*, *5*(2), e12305-e12305. https://doi.org/10.54021/seesv5n2-809

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

Chen, G., Liu, M., & Chen, J. (2020). Frequency-temporal-logic-based bearing fault diagnosis and fault interpretation using Bayesian optimization with Bayesian neural networks. *Mechanical Systems and Signal Processing*, *145*, 106951. https://doi.org/10.1016/j.ymssp.2020.106951

Chen, H.-Y., & Lee, C.-H. (2020). Vibration signals analysis by explainable artificial intelligence (XAI) approach: Application on bearing faults diagnosis. *IEEE Access*, *8*, 134246-134256. https://doi.org/10.1109/ACCESS.2020.3006491

Chen, Z., Wang, H., Zhou, Y., Yang, Y., & Liu, Y. (2024). A multi-scale feature extraction and fusion method for diagnosing bearing faults. *Journal of Dynamics, Monitoring and Diagnostics*, *3*(4), 268-278. https://doi.org/10.37965/jdmd.2024.560

Dai, W., Mo, Z., Luo, C., Jiang, J., Zhang, H., & Miao, Q. (2020). Fault diagnosis of rotating machinery based on deep reinforcement learning and reciprocal of smoothness index. *IEEE Sensors Journal*, *20*(15), 8307-8315. https://doi.org/10.1109/jsen.2020.2970747

Duong, B. P., & Kim, J.-M. (2018). Non-mutually exclusive deep neural network classifier for combined modes of bearing fault diagnosis. *Sensors*, *18*(4), 1129. https://doi.org/10.3390/s18041129

Eang, C., & Lee, S. (2024). Predictive maintenance and fault detection for motor drive control systems in industrial robots using CNN-RNN-Based Observers. *Sensors*, *25*(1), 25. https://doi.org/10.3390/s25010025

Eren, L., Ince, T., & Kiranyaz, S. (2019). A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *Journal of Signal Processing Systems*, *91*(2), 179-189. https://doi.org/10.1007/s11265-018-1378-3

Fan, F., Xiong, J., & Wang, G. (2020). Universal approximation with quadratic deep networks. *Neural Networks*, *124*, 383-392. https://doi.org/10.1016/j.neunet.2020.01.007

Gunerkar, R. S., Jalan, A. K., & Belgamwar, S. U. (2019). Fault diagnosis of rolling element bearing based on artificial neural network. *Journal of Mechanical Science and Technology*, *33*, 505-511. https://doi.org/10.1007/s12206-019-0103-x

Hu, J., Shen, L., & Sun, G. (2018). *Squeeze-and-excitation networks*. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Jalayer, R., Jalayer, M., Mor, A., Orsenigo, C., & Vercellis, C. (2024). Evaluating deep learning models for fault diagnosis of a rotating machinery with epistemic and aleatoric uncertainty. *arXiv preprint arXiv:2412.18980*. https://doi.org/10.48550/ARXIV.2412.18980

Liao, J.-X., Dong, H.-C., Sun, Z.-Q., Sun, J., Zhang, S., & Fan, F.-L. (2023). Attention-embedded quadratic network (qttention) for effective and interpretable bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1-13. https://doi.org/10.1109/TIM.2023.3259031

Liao, Y., Li, H., Zhang, L., Yan, Q., Yang, Y., & Li, B. (2023). *Anomaly transformer for vibration data analysis: Case study on CWRU Dataset*. Paper presented at the 2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), Chengdu, China: IEEE. https://doi.org/10.1109/ICICML60161.2023.10424762.

Lu, Q., Shen, X., Wang, X., Li, M., Li, J., & Zhang, M. (2021). Fault diagnosis of rolling bearing based on improved VMD and KNN. *Mathematical Problems in Engineering*, *2021*(1), 2530315. https://doi.org/10.1155/2021/2530315

Ozcan, I. H., Devecioglu, O. C., Ince, T., Eren, L., & Askar, M. (2022). Enhanced bearing fault detection using multichannel, multilevel 1D CNN classifier. *Electrical Engineering*, *104*(2), 435-447. https://doi.org/10.1007/s00202-021-01309-2

Pouyet, E., Rohani, N., Katsaggelos, A. K., Cossairt, O., & Walton, M. (2018). Innovative data reduction and visualization strategy for hyperspectral imaging datasets using t-SNE approach. *Pure and Applied Chemistry*, *90*(3), 493-506. https://doi.org/10.1515/pac-2017-0907

Qiu, S., Cui, X., Ping, Z., Shan, N., Li, Z., Bao, X., & Xu, X. (2023). Deep learning techniques in intelligent fault diagnosis and prognosis for industrial systems: A review. *Sensors*, *23*(3), 1305. https://doi.org/10.3390/s23031305

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 2: 2132-2146, 2025
DOI: 10.55214/25768484.v9i2.5050

Rai, A., & Upadhyay, S. H. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*, *96*, 289-306. https://doi.org/10.1016/j.triboint.2015.12.037

Rosa, R. K., Braga, D., & Silva, D. (2024). Benchmarking deep learning models for bearing fault diagnosis using the CWRU dataset: A multi-label approach. *arXiv preprint arXiv:2407.14625*. https://doi.org/10.48550/ARXIV.2407.14625

Shang, J., Chen, M., Ji, H., Zhou, D., Zhang, H., & Li, M. (2017). Dominant trend based logistic regression for fault diagnosis in nonstationary processes. *Control Engineering Practice*, *66*, 156-168. https://doi.org/10.1016/j.conengprac.2017.06.011

Shenfield, A., & Howarth, M. (2020). A novel deep learning model for the detection and identification of rolling element-bearing faults. *Sensors*, *20*(18), 5112. https://doi.org/10.3390/s20185112

Wan, L., Chen, Y., Li, H., & Li, C. (2020). Rolling-element bearing fault diagnosis using improved LeNet-5 network. *Sensors*, *20*(6), 1693. https://doi.org/10.3390/s20061693

Wang, H., Liu, Z., Peng, D., & Qin, Y. (2019). Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis. *IEEE Transactions on Industrial Informatics*, *16*(9), 5735-5745. https://doi.org/10.1109/TII.2019.2955540

Wang, H., Xu, J., Yan, R., Sun, C., & Chen, X. (2020). Intelligent bearing fault diagnosis using multi-head attention-based CNN. *Procedia Manufacturing*, *49*, 112-118. https://doi.org/10.1016/j.promfg.2020.07.005

Wang, M., Chen, Y., Zhang, X., Chau, T. K., Ching Iu, H. H., Fernando, T., . . . Ma, M. (2021). Roller bearing fault diagnosis based on integrated fault feature and SVM. *Journal of Vibration Engineering & Technologies*, *10*(3), 853–862. https://doi.org/10.1007/s42417-021-00414-7

Wang, T., Liang, M., Li, J., & Cheng, W. (2014). Rolling element bearing fault diagnosis via fault characteristic order (FCO) analysis. *Mechanical Systems and Signal Processing*, *45*(1), 139-153. https://doi.org/10.1016/j.ymssp.2013.11.011

Wang, Y., Ding, X., Zeng, Q., Wang, L., & Shao, Y. (2020). Intelligent rolling bearing fault diagnosis via vision ConvNet. *IEEE Sensors Journal*, *21*(5), 6600-6609. https://doi.org/10.1109/JSEN.2020.3042182

Xu, G., Liu, M., Jiang, Z., Söffker, D., & Shen, W. (2019). Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. *Sensors*, *19*(5), 1088. https://doi.org/10.3390/s19051088

Yoo, Y., Jo, H., & Ban, S.-W. (2023). Lite and efficient deep learning model for bearing fault diagnosis using the CWRU dataset. *Sensors*, *23*(6), 3157. https://doi.org/10.3390/s23063157

Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, *17*(2), 425. https://doi.org/10.3390/s17020425

Zhao, S., Zhu, Z., & He, J. (2024). *Research and optimization of feature extraction algorithms based on bearing vibration signals.* Paper presented at the International Conference on Cloud Computing and Communication Engineering (CCCE 2024), Y. Yue and H. Xing, Eds., Nanjing, China: SPIE. https://doi.org/10.1117/12.3049948.

Zhong, X., Wang, F., & Ban, H. (2022). Development of a plug-and-play anti-noise module for fault diagnosis of rotating machines in nuclear power plants. *Progress in Nuclear Energy*, *151*, 104344. https://doi.org/10.1016/j.pnucene.2022.104344