sciendo

POLISH MARITIME RESEARCH 3 (119) 2023 Vol. 30; pp. 132-141
10.2478/pomr-2023-0046

FAULT DIAGNOSIS OF BEARINGS BASED ON SSWT, BAYES OPTIMISATION AND CNN

Guohua Yan [1]
Yihuai Hu [2] *
Qingguo Shi [1]
[1] Department of Marine Engineering, School of Merchant Shipping, Shanghai Maritime University, China
[2] Department of Turbine Engineering, College of Merchant Marine, Shanghai Maritime University, China

* Corresponding author: *yhhu@shmtu.edu.cn (Yihuai Hu)*

ABSTRACT

Bearings are important components of rotating machinery and transmission systems, and are often damaged by wear, overload and shocks. Due to the low resolution of traditional time-frequency analysis for the diagnosis of bearing faults, a synchrosqueezed wavelet transform (SSWT) is proposed to improve the resolution. An improved convolutional neural network fault diagnosis model is proposed in this paper, and a Bayesian optimisation method is applied to automatically adjust the structure and hyperparameters of the model to improve the accuracy of bearing fault diagnosis. Experimental results from the accelerated life testing of bearings show that the proposed method is able to accurately identify various types of bearing fault and the different status of these faults under complex running conditions, while achieving very good generalisation ability.

Keywords: Fault diagnosis; Bearing; PMSM; Bayesian optimisation; CNN

INTRODUCTION

As bearings support the rotation of rotating bodies, they are important components of rotating machinery and transmission systems, and are widely used in industrial production and transportation. In harsh working environments, such as in ocean-going ships, bearings work at high temperatures and humidity, under corrosion from salt and alkalis, and are overloaded for long periods; they are therefore subject to wear and impact over the long term, which makes them extremely prone to various types of failure. In some types of equipment, bearings are the components that are most prone to failure. A statistical analysis has shown that the failure rate of bearings accounts for 30–40% of total motor failures [1], meaning that fault diagnosis of bearings could reduce or even prevent accidents, with very significant effects in terms of ensuring the safety of personnel and equipment.

Over the past few decades, researchers and engineers have developed a variety of methods for the diagnosis of bearing faults, which can be divided into analytical model-based, signal-based, and data-driven approaches [2]. Analytical model-based methods require a very good understanding of the operating mechanism of the motor and an accurate mathematical model, and diagnose the fault by analysing the difference between the model output and the actual output [3]. This method has poor flexibility and capability, and cannot be adapted to complex real-world applications. Signal-based methods do not require a complete model, and extract the features of the fault on the device from the time domain, frequency domain, or time-frequency domain [4]. Machine learning algorithms such as support vector machine (SVM) [5], an error backpropagation neural network (BPNN) [6] or K-means [7] can also be used to diagnose a fault; however, each feature extraction algorithm has certain limitations, and the extraction determines the final

result of diagnosis. Data-driven methods rely on big datasets, and can establish a mapping relationship between signals and faults from a large amount of historical data. They are very suitable for fault diagnosis in noisy environments.

Unlike traditional machine learning algorithms such as SVM and BPNN, deep learning algorithms with multiple nonlinear layers have a multi-layer feature learning capability, and can extract more high-quality feature information from a large amount of data; this means they can solve the problems inherent to traditional machine learning, and they are therefore widely used for bearing fault diagnosis. Deep autoencoders, deep belief networks, recurrent neural networks and convolutional neural networks (CNNs) are the most commonly used deep learning models for fault diagnosis [8]. Of these, the CNN is one of the most important, and has become a leading and fast-growing architecture for fault diagnosis. Zhang et al. [9] proposed a one-dimensional CNN algorithm for fault diagnosis from rotating machinery based on residual learning, which was shown to have higher accuracy than other commonly used deep learning fault diagnosis models. Pan et al. [4] proposed a combination of a one-dimensional CNN and an LSTM for bearing fault diagnosis, which made full use of the powerful feature extraction capability of the CNN and the good temporal modelling property of LSTM. Huang et al. [10] proposed a one-dimensional deep decoupled CNN fault diagnosis model that identified single and compound faults via feature learning and decoupled classification. However, the power of the CNN lies in its image processing capability, and there are many fault diagnosis methods that use two-dimensional CNNs. Time-frequency domain signal preprocessing methods are often used in combination with two-dimensional CNNs: this method uses time-frequency domain signal processing techniques to generate time-frequency spectra, and then exploits the powerful image processing capability of CNN to diagnose motor bearing faults. The wavelet transform (WT) method is a commonly used time-frequency analysis method. The WT time-frequency spectrum is based on wavelets, with the addition of time and scale variables, and can effectively handle nonstationary nonlinear signals. Ding et al. [11] proposed a multi-scale bearing fault feature extraction method with energy fluctuations using a wavelet packet energy map and a deep convolutional network. Verstraete et al. [12] used a WT, a short-time Fourier transform and a Hilbert-Huang transform (HHT) of the time-frequency spectrum as the input to a CNN to achieve automatic feature extraction and classification.

Although a WT has a high time-frequency resolution, it is still affected by the Heisenberg uncertainty principle, and the time-frequency resolution is somewhat limited, which makes the time-frequency spectrum blurred [13] and reduces the effectiveness of feature extraction in practical environments. In order to accurately describe nonstationary nonlinear vibration signals, many new methods have been proposed that improve on traditional WT methods, such as the synchrosqueezed WT (SSWT) [14] and the dual-core noise-reducing simultaneous squeezed WT [15]. Of these, SSWT is the most traditional, with good noise immunity and time-frequency resolution, and high robustness.

A CNN is a multi-layer feed-forward neural network, and a deep-learning-based CNN fault diagnosis model typically consists of two parts: filtering and classification. The filter stage is mainly used to extract the features of the input data, and includes a convolution layer, a batch normalisation layer, an active layer and a pooling layer. The classification module processes and classifies the extracted features, and includes a dropout layer and a fully connected layer. The training efficiency of a CNN is improved by reducing the number of parameters for model training through a parameter sharing mechanism and the use of sparse connections between the filtering and classification levels [16]. A CNN has a complex structure with many hyperparameters, and there is no mature theory to guide the design of the network structure; a trial-and-error method is generally used, and the results of experience and training are combined in an effort to obtain the right network architecture and hyperparameters. This requires numerous attempts and a considerable amount of time, and therefore poses a great challenge. Many hyperparameter optimisation methods have been developed, such as grid search, stochastic search, dynamic resource allocation, and Bayesian optimisation. Of these, Bayesian optimisation is notable as a classical adaptive optimisation method that takes the current optimal hyperparameters as a reference and predicts the next combination that will offer the greatest benefit.

To solve problems such as a blurred spectrum, a high level of noise and the difficulty of CNN design for traditional WT, a SSWT-Bayes-CNN fault diagnosis model is proposed in this paper. First, the SSWT method is used to increase the accuracy of representation of nonlinear, non-smooth vibration signals and to reduce noise interference in the signal. The SSWT time-frequency spectrum is used as input to the CNN, and the Bayesian optimisation algorithm is applied to achieve automatic optimisation of network hyperparameters, thereby reducing the dependence on network design experience. This process finally creates a high-performance CNN architecture, and can achieve rapid deployment of a bearing fault diagnosis model.

FAULT DIAGNOSIS METHOD BASED ON SSWT, BAYES OPTIMISATION AND CNN

SSWT

SSWT is a new mathematical transform tool that was developed on the basis of the WT, in a similar way to empirical mode decomposition (EMD) [17]. The WT can be seen as the result of a folded product of the signal with the mother wavelet after a telescopic translation, which forms the time-frequency spectrum and completes the sparse representation of the signal. The SSWT transform of the time signal $f(t)$ is based on the wavelet transform as [17]:

$$W_S(a, b)=<f(t), \varphi_{a,b}(t)>=\frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} s(t)\varphi^*(\frac{t-b}{a})dt \quad (1)$$

where $W_S(a, b)$ are the wavelet coefficients, and $<f(t), \varphi_{a,b}(t)>$ denotes the wavelet transform of the vibration signal $f(t)$. $\varphi(a, b)$

is the wavelet function, which is obtained from scaling and translating the wavelet basis function φ, i.e. [18]:

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}}\varphi\left(\frac{t-b}{a}\right) \quad (2)$$

where a, b are the scaling and scale factors, respectively, . The coefficient is normalised to the wavelet energy. According to Plancherel's theorem, Equation (1) can be written as an expression in the wave number domain, as follows [17]:

$$W_S(a,b) = \frac{A\sqrt{a}}{4\pi}\varphi^*(a\omega)e^{ib\omega} \quad (3)$$

Using the relationship between scale and frequency, the computed wavelet coefficients $W_S(a, b)$ can be mapped from the time scale plane to the time-frequency plane $(b, a) \rightarrow b, W_S(a, b)$. The coefficients of SSWT lie only in the frequency interval $[(\omega_l-\Delta\omega/2), (\omega_l+\Delta\omega/2)]$, centered at ω, while $\Delta\omega=\omega_l-\omega_{l-1}$. Since the scaling factor a and the scale factor b are discrete values, a scaling step $\Delta a_k=a_k-a_{k-1}$ is taken to isolate the discrete scale a_k [18], and then its wavelet decomposition is computed using Equation (3). The SSWT is [14]:

$$T_S(\omega_l, b)=(\Delta\omega)^{-1}\sum_{a_k:[W(a_k,b)-W_l]\le\Delta a_k} W_S(a_k, b)a_k^{-3/2}\Delta a_k \quad (4)$$

From Equation (4), it can be seen that the SSWT compresses the wavelet coefficients only in the frequency direction. This compression process preserves the phase information of the wavelet coefficients, and the simultaneous compression WT can be reconstructed with the inverse transform as [14]:

$$S(b) = Re[C_\Psi^{-1}\sum T_S(\omega_l, t)\,(\Delta\omega) \quad (5)$$

To illustrate the advantages of SSWT, a simple segment of the signal $f(t)$ was constructed with the following signal:

$$f(t) = 2sin(600t)+3sin(300t)+5sin(150t) \quad (6)$$

Noise with a signal-to-noise ratio (SNR) of 15 dB was added to the signal $f(t)$ to simulate a complex signal with nonlinear nonstationarity, and three mainstream time-frequency analysis methods Continuous Wavelet Transform (CWT), HHT and SSWT) were applied to the signal $f(t)$ for comparison. The results are shown in Fig. 1. CWT has a very vague time-frequency performance, whereas HHT can clearly express the main components of the signal, although the resolution is not high and the detection band is wide. Unlike the vague time-frequency information representation of CWT, SSWT overcomes the problem of low time-frequency resolution of the signal, and characterises the time-frequency relationship of the signal more clearly and accurately.

BAYESIAN OPTIMISATION

The Bayesian optimisation algorithm can predict the maximum of a function based on existing sample points when the equation of the function is unknown. Its probabilistic agent model uses Gaussian processes (GPs). In our deep learning model, the combination of hyperparameters to be optimised is assumed to be $x=\{x_1, x_2\ldots, x_n\}$. The objective function $f(x)$ of the Bayesian optimisation algorithm is the classification error of the trained deep learning model on the validation set. The classification of each dataset in the test set is considered an independent event with a certain probability of success, meaning that the number of datasets with classification errors follows a binomial distribution [19]:

$$f(t) = GP(\mu, k(x,x)) \quad (7)$$

where GP is the Gaussian distribution, μ is the mean, and $k(x,x)$ is the covariance function.

After $n-1$ iterations of the Bayesian optimisation algorithm on the dataset $D_{n-1}=\{(x_1,f(x_1)), (x_2,f(x_2)),\ldots,(x_{n-1},f(x_{n-1}))\}$, the next step is to predict the observation $f(x_n)$ at the point x_n. It is generally assumed that these n observations are samples of an n-dimensional Gaussian distribution, i.e.:

$$\left[\frac{f_{1:n-1}}{f_n}\right]\sim GP\left(\mu,\left[\frac{K,k^T}{k,k(x_n,x_t)}\right]\right) \quad (8)$$

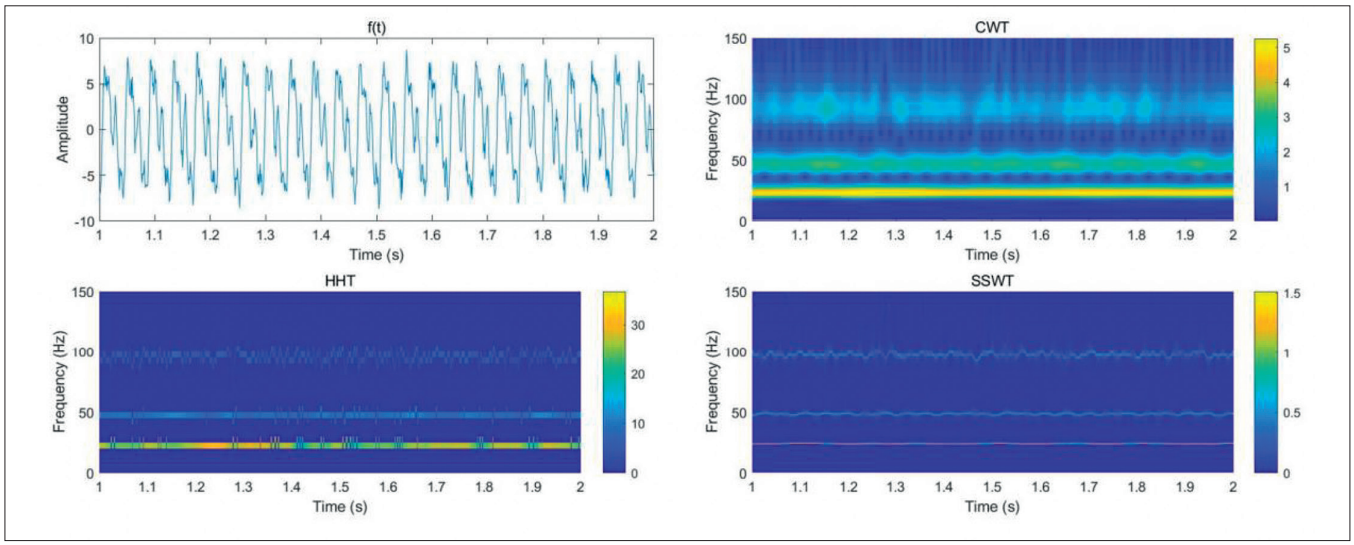


Fig. 1. SSWT and CWT time spectra

where $K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_{n-1}) \\ \vdots & \ddots & \vdots \\ k(x_{n-1}, x_1) & \cdots & k(x_{n-1}, x_{n-1}) \end{bmatrix}$,

$$k = \left[(x_n, x_1), (x_n, x_2), \ldots, (x_n, x_{n-1})\right].$$

This gives a Gaussian distribution for f_n:

$$P(f_n|D, x_n) = GP\left(\mu(x_n), \delta^2(x_n)\right) \tag{9}$$

where $u(x_i)=kK^{-1}f_{1:n-1}$ is the predicted mean, and $\delta^2(x_n)=k(x_n, x_n)-kK^{-1}k^T$ is the predicted covariance. $(x_n, f(x_n))$ is the final data item for the nth observation point.

In order to use as few samples as possible, to make the Gaussian distribution of f_n closer to the true distribution, and to reduce the number of samples taken from the sample space, the acquisition function is used to determine whether a sample can can improve the model. The larger the gain, the closer the modified Gaussian process will be to the true distribution of objective function. The expected improvement (EI) function can reduce the probability of local optimisation, and is therefore chosen as the acquisition function.

IMPROVED CNN MODEL

The current trend in CNN models is to increase the number of network layers and decrease the size of the convolutional kernels. Smaller convolutional kernels can obtain better local information, whereas deeper networks can obtain better global information. An improved CNN model is proposed as shown in Fig. 2. In Conv_1, the number of layers is indeterminate for information filtering, feature extraction and size compression of the SSWT time spectrum, where the number of layers is determined by the actual needs. Each convolutional layer in Conv_1 includes batch normalisation (BN) to speed up convergence of the network and prevent overfitting, and ReLU is used as the activation function. Conv_2 and Conv_3 have the same internal structure and number of layers as Conv_1; the only difference is that Conv_1 uses 16 convolutional kernels, whereas Conv_2 uses 32 and Conv_3 uses 64. The number of convolutional kernels increases with the depth of the network. The deeper the network, the smaller the computation, and the additional convolutional kernels can use these computational capabilities. In addition, the deeper the network, the more abstract the feature information that is extracted. An increase in the number of convolutional kernels allows the feature information previously learned by the network to be better combined, meaning that various features can be covered more comprehensively. The fully connected layer (FC) can obtain all the feature information extracted from the convolutional layer, but it suffers from too many parameters and redundancy, which tends to cause overfitting of the network. The addition of another FC layer can improve the overfitting situation, but makes the network more bloated. The Global Average Pooling (GAP) layer can integrate the high-dimensional information output from the convolutional layer, thus directly achieving a significant reduction in the number of feature parameters. This layer has no parameters, thereby avoiding overfitting, and has better robustness. The final classification layer applies a softmax function, which is used to output the probability of each fault.

BEARING FAULT DIAGNOSIS BASED ON BAYESIAN OPTIMISATION AND CNN

The hyperparameters of the CNN model have an important impact on the training process and model representation, and include the depth of the convolutional layer *n*, the momentum factor, the L2 regularisation, and the initial learning rate. Overfitting means that although the model has a very low error on the training set (original dataset), it has a very high error on the unseen test set (new dataset), i.e. the model has poor generalisation ability, and cannot be generalised effectively from the original dataset to another dataset. Underfitting means that the model cannot obtain a sufficiently low error on the training set. The larger the size of the L2 regularisation parameter, the stricter the constraint, and the easier it is to cause underfitting of the model. If the parameter is too small, the constraint is loose, and does not act as a constraint, which tends to produce overfitting [20]. An initial learning rate that is too large will result in a model that fails to converge and an unstable training process, whereas a rate that is too small will result in a model that converges particularly slowly or fails to learn. The momentum-based stochastic gradient descent method (SGDM) introduces a momentum factor that ameliorates the problem of traditional stochastic gradient descent (SGD) oscillations. When the network tends to converge in the middle and late stages of training, the network parameters oscillate back and forth around local minima, which also helps the network to jump out of local boundaries and find better network parameters. If the momentum factor is too high, it may cause the model to oscillate in the parameter space and fail to converge stably, whereas a momentum factor that is too small may cause the model to converge too slowly or even to fall into a local minimum. The hyperparameters of the Bayesian optimisation-based CNN bearing fault diagnosis model are the convolutional layer depth *n*, the momentum factor, the L2

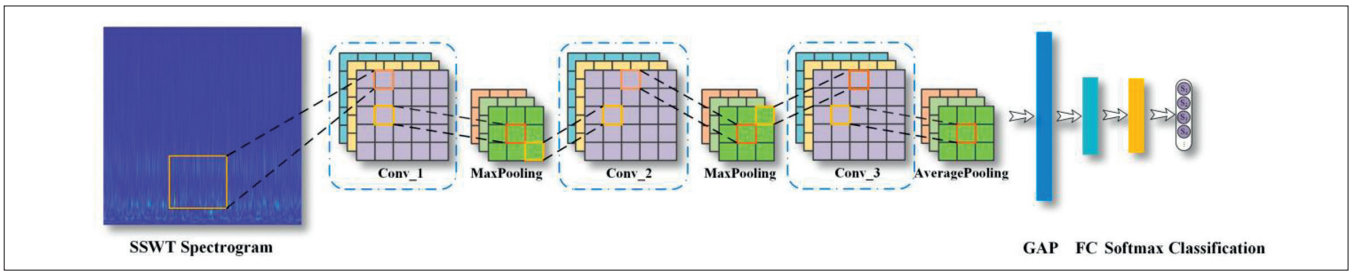


Fig. 2. Improved CNN model

regularisation, and the initial learning rate, as shown in Fig. 3. In general, when a supervised deep learning model is trained, the data are divided into training, validation and test sets. The results on the validation set can accurately reflect the current training status of the network, and can verify the generalisation ability of the model. The error on the validation set after completion of training is therefore used in this paper as the objective function for Bayesian optimisation, denoted as $f(x)$, and the optimisation objective is to minimise the error on the validation set.

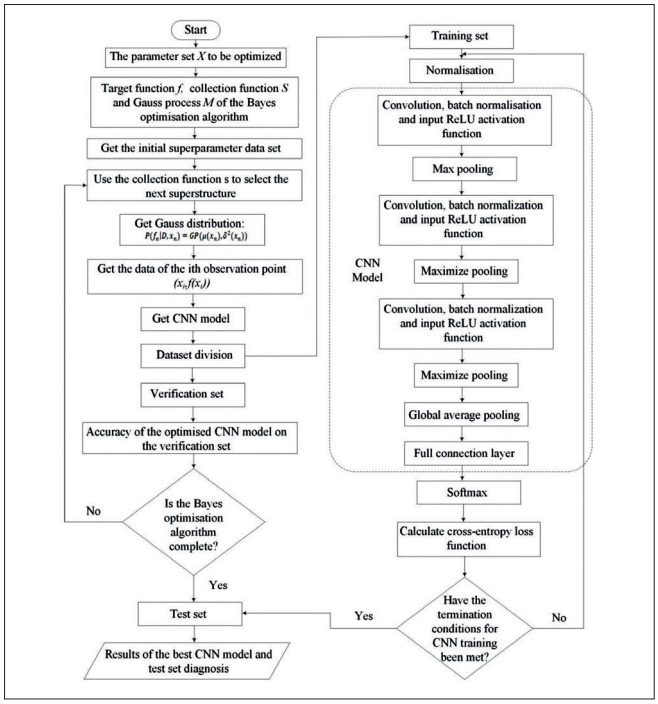
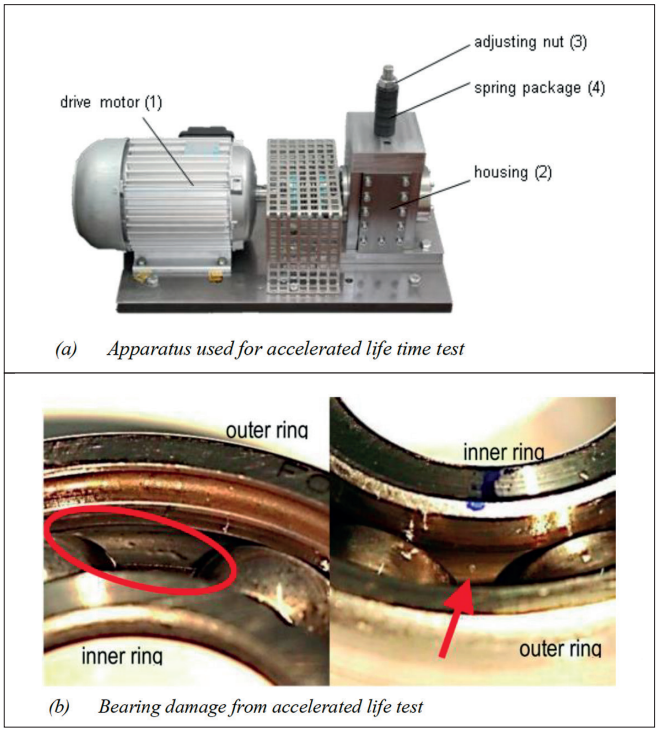


Fig.3. Fault diagnosis model based on Bayesian optimisation of a CNN

TEST ANALYSIS AND BEARING FAULT DIAGNOSIS

TEST DATA

The acquisition of high-quality bearing fault signals is still a difficult task, due to the limitations of data acquisition techniques, the time and cost required to obtain natural fault data, and other problems. Fortunately, some research institutions have published such datasets, including the Case Western Reserve University (CWRU) bearing dataset [21], the MFPT fault dataset [22], and the Paderborn dataset [23]. The CWRU dataset is one of the most commonly used for bearing fault diagnosis; however, all of the the bearings in this dataset are artificially damaged and operate under simple conditions, unlike actual bearings. The Paderborn dataset, provided by Paderborn University, Germany, contains a variety of fault types and severities; the working conditions of the bearings are diverse; both artificial and actual damage are considered; and the dataset is relatively comprehensive. Lessmeier et al. [23] carried out a detailed study using WPD and FFT to extract fault features, and compared the results from popular machine learning classification algorithms such as a neural network (NN), random forest (RF) and SVM, which gave good accuracy on both artificial and actual damage datasets. However, when a model is trained on an artificial damage dataset and tested on actual damage dataset, poor diagnostic results are generally obtained. This indicates that there are significant differences between artificial damage and actual faults, and that artificial faults cannot accurately represent real ones. Hence, in order to more closely match the failure situation in real industrial scenarios, only real failure data are selected from accelerated life tests, as shown in Fig. 4. The dataset uses 6,203 grooved ball bearings, including single and repeated damage to the inner ring, single and repeated damage to the outer ring, and compound damage to the inner and outer rings. Based on the length of the damage and the percentage size of the pitch circumference, the faults are divided into five levels indicating their severity, as shown in Table 1.



(a) Apparatus used for accelerated life time test

(b) Bearing damage from accelerated life test

Fig.3. Fault diagnosis model based on Bayesian optimisation of a CNN

Tab. 1. Bearing fault levels [21]

Damage level	Assigned percentage values	Limits for bearing 6203
1	0–2%	≤ 2 mm
2	2–5%	> 2 mm
3	5–15%	> 4.5 mm
4	15–35%	> 13.5 mm
5	>35%	> 31.5 mm

The sampling frequency of the vibration signal was 64 kHz, the sampling length was 4 s, and a total of 20 samples were considered for each fault. Detailed information on the samples is shown in Table 2.

Tab. 2. Operating conditions and status of test bearings

Bearing status	Working conditions	Damage level	Operating state
Health	• n = 700 r/min, M = 0.7 Nm • n = 1500 r/min, M = 0.1 Nm • n = 1500 r/min, M = 0.7 Nm	0	H
Damage to inner ring	• n = 700 r/min, M = 0.7 Nm • n = 1500 r/min, M = 0.1 Nm • n = 1500 r/min, M = 0.7 Nm	1	IR I
	• n = 700 r/min, M = 0.7 Nm • n = 1500 r/min, M = 0.1 Nm • n = 1500 r/min, M = 0.7 Nm	2	IR II
	• n = 700 r/min, M = 0.7 Nm • n = 1500 r/min, M = 0.1 Nm • n = 1500 r/min, M = 0.7 Nm	3	IR III
Damage to outer ring	• n = 700 r/min, M = 0.7 Nm • n = 1500 r/min, M = 0.1 Nm • n = 1500 r/min, M = 0.7 Nm	1	OR I
	• n = 700 r/min, M = 0.7 Nm • n = 1500 r/min, M = 0.1 Nm • n = 1500 r/min, M = 0.7 Nm	2	OR II
Damage to both inner and outer rings	• n = 700 r/min, M = 0.7 Nm • n = 1500 r/min, M = 0.1 Nm • n = 1500 r/min, M = 0.7 Nm	1	IR+OR I
	• n = 700 r/min, M = 0.7 Nm • n = 1500 r/min, M = 0.1 Nm • n = 1500 r/min, M = 0.7 Nm	2	IR+OR II
	• n = 700 r/min, M = 0.7 Nm • n = 1500 r/min, M = 0.1 Nm • n = 1500 r/min, M = 0.7 Nm	3	IR+OR III

SSWT PRE-PROCESSING

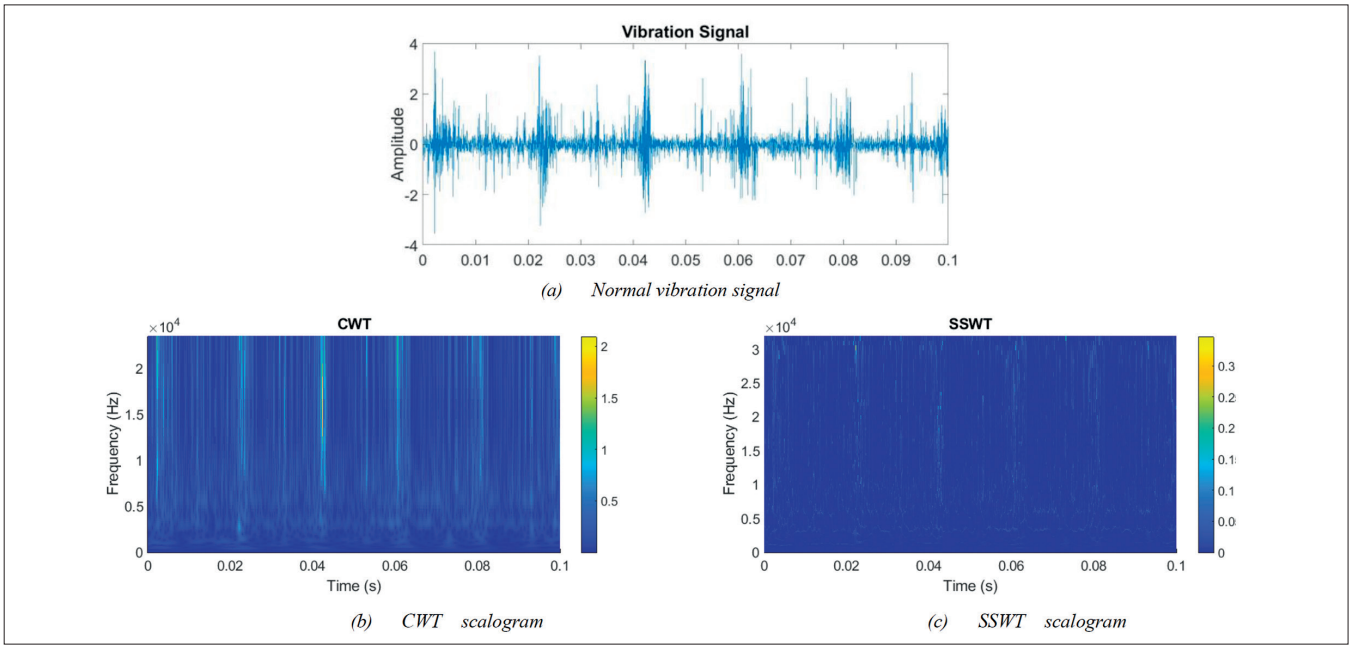
As can be seen from Table 2, there are nine possible operating states of the bearings under three different working conditions. For each working condition, the dataset contains 20 samples, meaning that it is very rich. Data of length 0.1 s are selected as samples. The original vibration data are sliced, with 9 working conditions and 2400 samples were obtained for each working condition, each sample length is 6400. Each sliced sample is processed via SSWT to obtain a time-frequency spectrum of size 227×227×3, which is used as the input to the network. Fig. 5 shows the CWT and SSWT time-frequency spectra of the real bearing vibration signal. It can be seen that the SSWT method greatly improves the time-frequency resolution of the signal, and can express the time-frequency relationship of the signal more accurately and clearly.

OPTIMISATION OF THE CNN HYPERPARAMETERS

The SSWT time data were divided into three groups: a training set, a validation set and a test set. The test set for each working condition was a 240 fixed group, and the training set and validation set were randomly assigned in the ratio of 4:1 in the remaining sample. The final numbers of samples in the training, validation and test sets for each working condition were 1728, 432 and 240, respectively. The datasets were input into the CNN fault diagnosis model with the objective of minimising the validation set error,, and a Bayesian algorithm is used to optimise the four hyper-parameters in the model to find the minimum value of the objective function $f(x)$, as shown in Table 3. Table 4 shows the main options for training the CNN fault diagnosis model. To reduce the possibility of overfitting in the training process, data enhancement techniques were applied to the training set, including random scaling, rotation and panning of the images.

Tab. 3. Optimised hyperparameters and ranges

Hyperparameters	Search space	Data type
Section depth	1 – 3	int
Initial learning rate	5×10^{-3} – 1	log
Momentum	0.8 – 0.98	log
L2 regularisation	1×10^{-10} – 0.01	log



(a) Normal vibration signal

(b) CWT scalogram

(c) SSWT scalogram

Fig. 5. CWT and SSWT time spectra for a normal bearing vibration signal

Tab. 4. Main options used for model training

Optimisation algorithm	SGDM	Shuffle	Every epoch
Learning rate drop factor	0.5	Validation frequency	Every epoch
Learning rate drop period	10	MaxEpochs	50
Learn rate schedule	Piecewise	Minibatch size	128

The number of Bayesian optimisations was set to 35, to train the CNN fault diagnosis model using the SGDM optimiser, the optimiser stores the hyperparameters for each optimisation and the accuracy of the validation set. The prediction of the current objective function was calculated by a Gaussian process, nd the next set of hyperparameters are selected using the gain function. The CNN model was retrained using new hyperparameters for ,Fig. 6 shows the minimum objective function versus the number of Bayesian optimisations, and it can be seen that the minimum objective function is obtained after the 23rd optimisation, which represents the optimal combination of hyperparameters. The depth of the CNN network is three, the initial learning rate is 0.022158, the momentum for the stochastic gradient descent is 0.89684, and the coefficient of L2 regularisation is 2.2395×10^{-7}. Fig. 7 shows the accuracy on the validation set after each optimisation. The accuracy on the validation set after the fourth optimisation is much lower than at the other points, which is due to the fact that the hyperparameters chosen for this optimisation trap the network in a local minimum, thus resulting in low accuracy.

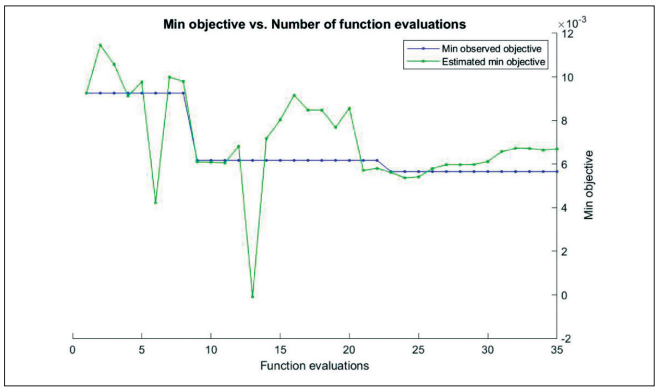


Fig. 6. Predicted values, observed values and evaluation times for the minimum objective function

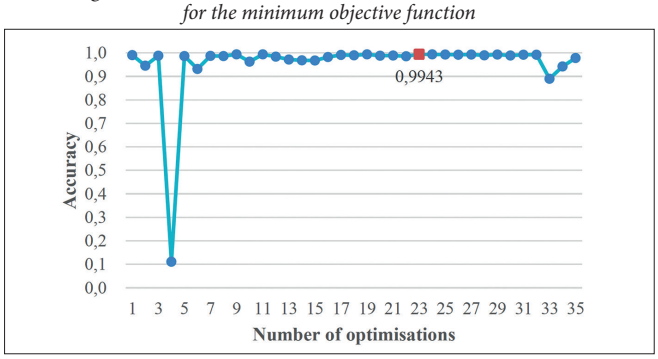


Fig. 7. Accuracy on the verification set vs. number of evaluations

The test set was used to verify the trained CNN fault diagnosis model. The accuracy of the test set can be used to measure the performance of the model, and is expressed as:

$$Accuracy = \frac{n}{N} \times 100\% \quad (10)$$

where n is the number of correctly classified samples, and N is the total number of samples in the test set. Fig. 8 shows the test results in the form of a confusion matrix, and it can be seen that the proposed method obtained an accuracy of 99.40%. The confusion matrix shows that the proposed model can distinguish between normal operation and a fault on the bearing with 100% accuracy; the locations of the bearing faults are almost all accurately classified, and the main problem with misclassification is associated with the severity of the damage. The highest diagnostic error rates are seen for the fault statuses of the inner rings II and III. This is because the faults are located in the inner ring, with high similarity between the two characteristics, which increases the difficulty of classification. Multiple failure types, severities and operating conditions mean that the bearing operating characteristics have more diversity and similarity, but this also increases the difficulty of classification. Despite the complex and diverse working conditions of the bearings, the proposed fault diagnosis model achieves very good results.

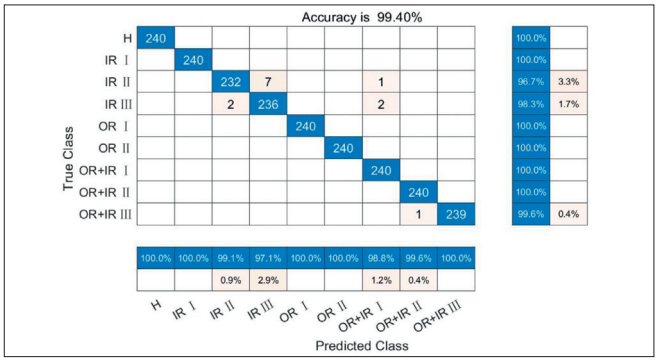


Fig. 8. Confusion matrix for the classification results from the optimal CNN model on the test set

In the testing process, the classification of each image in the test set is considered an independent event with a certain probability of success, i.e. the number of misclassified images follows a binomial distribution, where there are only two classification outcomes for each image: correct or incorrect. The Wald test for CNN fault diagnosis models consists of two steps, as follows:

(i) Calculate the standard error of for the test set:

$$TestErrorSE = \sqrt{\frac{Accuracy(1-accuracy)}{N}} \quad (11)$$

(ii) Calculate the 95% confidence interval for the standard error:

$$TestError95CI = [1 - Accuracy - 1.96^{*}\ TestErrorSE,$$

$$1 - Accuracy + 1.96^{*}\ TestErrorSE] \quad (12)$$

where *Accuracy* is the accuracy on the test set obtained from Equation (10). Both the standard error for the test set and the 95% confidence interval of the standard error reflect the performance of the proposed CNN fault diagnosis model, and

the standard variance reflects the degree of dispersion of the results for the test set. The larger the standard variance, the more samples in the test set are misclassified, whereas the smaller the 95% confidence interval, the more samples in the test set are correctly classified. The result for *TestErrorSE* is 0.0017, and the *TestError95CI* interval is [0.0028, 0.0093]; both of these are very small, indicating that the CNN fault diagnosis model proposed in this paper has very good performance.

GENERALISATION AND ROBUSTNESS TESTING OF MODELS IN NOISY ENVIRONMENTS

A bearing operates in a variety of environments under actual operating conditions, and the collected signals contain some random interference noise, which affects the effectiveness of fault diagnosis. The SNR is the ratio of the average power of the effective signal to the average power of the noise, as follows:

$$SNR = 10log_{10}\frac{P_{Signal}}{P_{Noise}} \quad (13)$$

where P_{Signal} is the energy of effective signal and P_{Noise} is the energy of the noise. The smaller the SNR, the higher the level of noise in the signal.

Gaussian white noise signals with values for the SNR of 0–80 dB at 10 dB intervals were added to the original vibration signals, and nine levels of noise intensity were applied as a way to simulate the acquisition of bearing vibration signals under different actual environments. These values covered most operating environments. The bearing vibration signals under different noise conditions were processed by SSWT. The CNN model and the hyperparameters were trained and tested, and the results are shown in Fig. 9. It can be seen that the noise intensity affects the accuracy of fault diagnosis; however, the accuracy of the proposed model is greater than 99% for the low-noise case, 98% in the medium-noise case, and higher than 90% in the high-noise case. It can therefore be concluded that the proposed fault diagnosis model gives good diagnostic results even in a high-noise environment.

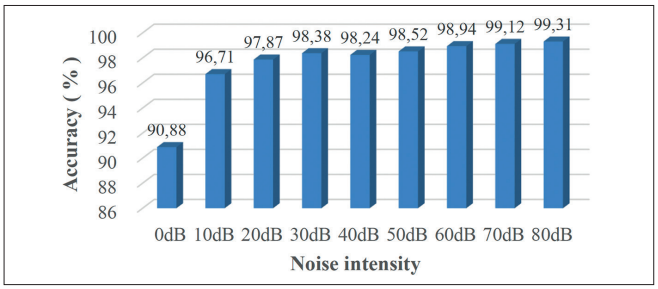


Fig. 9. Accuracy of CNN models trained and tested with different levels of noise

COMPARISON WITH OTHER FAULT DIAGNOSIS METHODS

CWT can closely model the detailed characteristics of vibration signals, and is widely used in the diagnosis of bearing faults. A Fourier simultaneous squeeze transform (FSST) is also a method for improving the time-frequency resolution. In a similar way to SSWT, FSST improves the time-frequency resolution of STFT by rearranging the time-frequency coefficients with a simultaneous squeeze operator. The raw vibration signal from the bearing was processed with both CWT and FSST, and the same CNN network model and hyperparameters as mentioned above were used for training and testing. The results are shown in Fig. 10, and it can be seen that SSWT gives the best diagnostic performance.

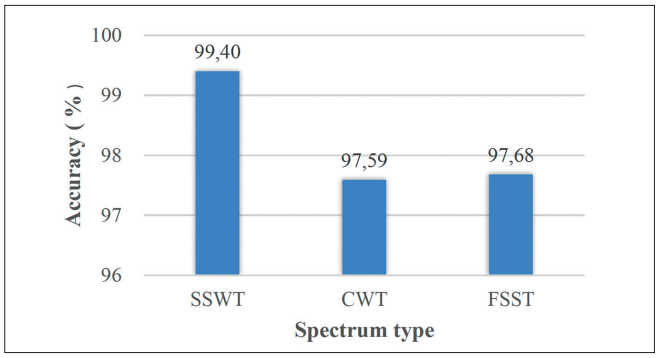


Fig. 10. Accuracy of the model using three time spectra

AlexNet was the first modern deep convolutional network model, and won the ImageNet image classification competition in 2012 [25]. During the recent boom in deep learning, other classical deep convolutional network models such as GoogLeNet and ResNet were developed. Using the original vibration signal and the spectrogram after SSWT processing under different noise signals, the classical models mentioned above were used for training and testing, and the results are shown in Fig. 11. It can be seen that the model proposed in this paper gave the best diagnostic performance under all levels of noise intensity.

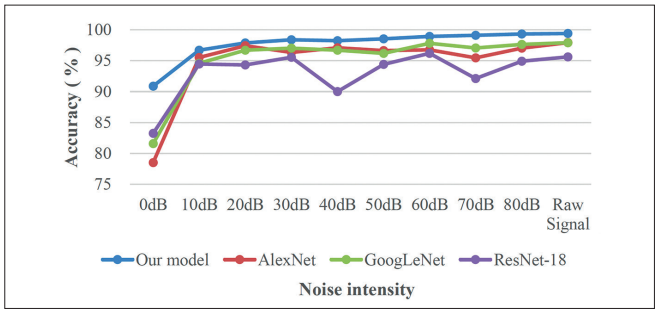


Fig. 11. Accuracy of different models under varying noise conditions

CONCLUSION

An adaptive bearing fault diagnosis method (SSWT-Bayes-CNN) has been proposed in this paper with the aim of adapting to the changable characteristics of bearing fault signals under complex operating conditions. The SSWT method is used to improve the resolution of the time-frequency spectrum, in order to highlight the most important information in the complex signal. To address the difficulty of selecting an CNN structure and appropriate hyperparameters, an improved CNN model was proposed, and a Bayesian optimisation method was applied to determine the optimal structure and hyperparameters for

the CNN, which provided a solution to the selection problem of multiple hyperparameter coupling. Experimental results confirmed that the proposed bearing fault diagnosis method could effectively diagnose different types and severities of bearing faults under complex operating conditions, with very high accuracy and very good generalisation ability. The use of accelerated life test data also provided a reference for the application of our model to practical fault diagnosis.

ACKNOWLEDGEMENTS

This work was supported by the Science & Technology Commission of Shanghai Municipality and Shanghai Engineering Research Centre for Intelligent Ship Maintenance and Energy Efficiency under Grant 20DZ2252300.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

1. S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020, doi: 10.1109/ACCESS.2020.2972859.

2. J. A. Reyes-Malanche, F. J. Villalobos-Pina, E. Cabal-Yepez, R. Alvarez-Salas, and C. Rodriguez-Donate, "Open-circuit fault diagnosis in power inverters through currents analysis in time domain," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021, Art no. 3517512, doi: 10.1109/TIM.2021.3082325.

3. X. Chen, P. Qin, Y. Chen, J. Zhao, W. Li, Y. Mao, and T. Zhao, "Inter-turn short circuit fault diagnosis of PMSM," *Electronics*, vol. 11, no. 10, p. 1576, 2022, https://doi.org/10.3390/electronics11101576.

4. H. Pan, X. He, S. Tang, and F. Meng, "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM," 2018, bearing fault diagnosis; CNN; LSTM vol. 64, no. 7–8, p. 10, 2018.

5. S. Liang, Y. Chen, H. Liang, and X. Li, "Sparse representation and SVM diagnosis method for inter-turn short-circuit fault in PMSM," *Applied Sciences*, vol. 9, no. 2, p. 224, 2019, https://doi.org/10.3390/app9020224.

6. Z. Zhao, Q. Xu, and M. Jia, "Improved shuffled frog leaping algorithm-based BP neural network and its application in bearing early fault diagnosis," *Neural Computing and Applications*, vol. 27, no. 2, pp. 375–385, 2016.

7. L.-K. Chang, S.-H. Wang, and M.-C. Tsai, "Demagnetization fault diagnosis of a PMSM using auto-encoder and K-means clustering," *Energies*, vol. 13, no. 17, doi: 10.3390/en13174467.

8. J. Jiao, M. Zhao, J. Lin, and K. Liang, "A comprehensive review on convolutional neural network in machine fault diagnosis," *Neurocomputing*, vol. 417, pp. 36–63, 2020.

9. W. Zhang, X. Li, and Q. Ding, "Deep residual learning-based fault diagnosis method for rotating machinery," *ISA Transactions*, vol. 95, pp. 295–305, 2019.

10. R. Huang, Y. Liao, S. Zhang, and W. Li, "Deep decoupling convolutional neural network for intelligent compound fault diagnosis," *IEEE Access*, vol. 7, pp. 1848–1858, 2019, doi: 10.1109/ACCESS.2018.2886343.

11. X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 1926–1935, Aug. 2017, doi: 10.1109/TIM.2017.2674738.

12. D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock and Vibration*, vol. 2017, p. 5067651, 2017.

13. Z. Shi, X. Yang, Y. Li, and G. Yu, "Wavelet-based synchroextracting transform: An effective TFA tool for machinery fault diagnosis," *Control Engineering Practice*, vol. 114, p. 104884, 2021.

14. C. Su, et al., "Damage assessments of composite under the environment with strong noise based on synchrosqueezing wavelet transform and stack autoencoder algorithm," *Measurement*, vol. 156, p. 107587, 2020.

15. J. Yuan, Z. Yao, Q. Zhao, Y. Xu, C. Li, and H. Jiang, "Dual-core denoised synchrosqueezing wavelet transform for gear fault detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021, Art no. 3521611, doi: 10.1109/TIM.2021.3094838.

16. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

17. G. W. Chang, Y.-L. Lin, Y.-J. Liu, G. H. Sun, and J. T. Yu, "A hybrid approach for time-varying harmonic and interharmonic detection using synchrosqueezing wavelet transform," *Applied Sciences*, vol. 11, no. 2, p. 752, 2021.

18. H. Wang and D.-Y. Yeung, "Towards Bayesian deep learning: A framework and some existing methods," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 12, pp. 3395–3408, Dec. 2016, doi: 10.1109/TKDE.2016.2606428.

19. M. Jiaocheng, S. Jinan, Z. Xin, and Z. Peng, "Bayes-DCGRU with Bayesian optimization for rolling bearing fault diagnosis," *Applied Intelligence*, vol. 52, no. 10, pp. 11172–11183, 2022.

20. W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64–65, pp. 100–131, 2015.

21. W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," Mechanical Systems and Signal Processing, vol. 64-65, pp. 100-131, 2015.

22. USA. The Vibration Institute, Condition Based Maintenance Fault Database for Testing of Diagnostic and Prognostics Algorithms. [Online]. https://www.mfpt.org/fault-data-sets..

23. C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," In: Editor. Pub Place; 2016. pp. 5–8.

24. Germany. University of Paderborn, Department of Design and Drive Technology, Condition Monitoring (CM) Experimental Bearing Dataset Based on Vibration and Motor Current Signals. [Online]. https://mb.uni-paderborn.de/kat/forschung/kat-datacenter/bearing-datacenter/data-sets-and-download

25. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun*. ACM, vol. 60, no. 6, pp. 84–90, 2017.