Pertanika J. Sci. & Technol. 33 (1): 261 - 282 (2025)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Classification of Bearing Degradation Stage Based on Automatic Label Assignment and Multi-scale Channel-attention Network

Xiuyu Li[1,2] and Shirley Johnathan Tanjong[1]*

[1]*Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, 94300, Sarawak, Malaysia*
[2]*Intelligent Manufacturing and Automobile Academy, Anhui Business and Technology College, 231131, HeFei, China*

ABSTRACT

Predicting bearing degradation is crucial for precise maintenance. However, accurately predicting the degradation stages of bearings to achieve appropriate maintenance has always been challenging. To address this problem, we propose a network architecture based on automatic label assignment called FAEK and a multi-scale channel-attention classification (MCC) prediction model to predict the degradation stage of bearings at a given time. Our method achieved outstanding performance on the FEMTO dataset with an accuracy of 0.9665. This approach provides an efficient and reliable solution for the predictive maintenance of bearings.

Keywords: Automatic label assignment, bearing degradation prediction, classification prediction model, deep learning, predictive maintenance

INTRODUCTION

Today, with the rapid development of technology and the continuous advancement of industrialisation, the reliability and performance of equipment and systems have become important focal points across various fields. Prognostics and health management (PHM) (Xia & Xi, 2019), an advanced technology and approach, offers an effective means for predicting equipment failure, optimising maintenance, and enhancing reliability through real-time monitoring, diagnostics, and prognostics of equipment health.

PHM utilises operational data from equipment to predict health conditions and enable precision maintenance. Compared with traditional methods involving routine inspections and regular checkups, PHM provides an intelligent maintenance

ARTICLE INFO

Article history:
Received: 18 May 2024
Accepted: 7 October 2024
Published: 27 January 2025

DOI: https://doi.org/10.47836/pjst.33.1.12

E-mail addresses:
21010282@siswa.unimas.my (Xiuyu Li)
jtshirley@unimas.my (Shirley Johnathan Tanjong)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

approach. In conventional maintenance, determining the optimal timing for maintenance can be challenging; however, with PHM, the stage of the life cycle of the equipment can be predicted, enabling the implementation of appropriate maintenance strategies. This precise maintenance approach significantly enhances equipment maintenance's efficiency and cost-effectiveness while extending its operational lifespan. PHM technology offers equipment managers a more intelligent and efficient maintenance approach to managing large-scale equipment and complex operational requirements.

This paper focuses on bearing components, critical parts of many rotating devices. Their failure can result in equipment downtime and production interruptions, causing significant business losses. PHM technology can establish fault prediction models by collecting and analysing vibration data from bearings. Monitoring changes in the bearing's condition can predict its remaining lifespan. Currently, two main methods exist for predicting the remaining lifespan of bearings. First, the direct prediction method, which relies on machine learning models to directly predict a bearing's lifespan (Li et al., 2019; Wang et al., 2019; Ruan et al., 2023). By inputting feature data into the model, the model directly outputs the predicted value of the remaining lifespan. This approach requires selecting and extracting appropriate features and demands a high level of expertise in feature engineering. However, it cannot explain the bearing's degradation and failure mechanisms. Second, the health indicator method predicts the time when the bearing reaches its failure threshold by monitoring its degradation process (Wang et al., 2021; Ren et al., 2018; Sanakkayala et al., 2022). Subsequently, it calculates the remaining useful life. This approach requires the prior determination of suitable health indicators and thresholds, which relies on understanding the system's characteristics and degradation behaviour, making it challenging.

In the PHM of bearings, the regression problem of predicting the remaining life of a bearing is transformed into a classification problem. Rather than solely predicting the remaining life of the bearing, we are more interested in determining the current stage of the bearing’s lifecycle. It enables us to conduct more effective maintenance and management procedures. To achieve this, we can divide the entire lifespan of the bearing into different stages, and with the given bearing data, we can predict the current stage of the bearing life. It can aid in identifying appropriate maintenance strategies and operational measures, enabling us to implement suitable preventive maintenance or repair actions at different stages.

Based on this, this paper proposes a method for the early prediction of bearing degradation. This method combines a convolutional autoencoder segmentation network with a high-performance classifier to effectively differentiate and accurately classify bearing degradation processes. It provides more comprehensive and reliable support for the health assessment of bearings and maintenance decision-making. The contributions of this study are as follows:

- We propose a model based on convolutional autoencoders that can selectively distinguish different stages of bearing degradation. The network learns a low-dimensional representation of the input data and uses reconstruction to segment the data into different stages of degradation. By utilising a convolutional autoencoder segmentation network, we can extract crucial features from bearing vibration data and employ them for subsequent classification tasks.
- We propose a high-performance classifier for classifying and identifying the different stages of bearing degradation. To better adapt to the bearing degradation classification task, we extract both frequency and time-domain features based on the vibration data of the bearings used to train the classifier.
- The classifier architecture is based on multi-branch parallel convolution and attention mechanisms, enabling it to capture long-distance dependencies within sequential data and learn the correlations between features. Consequently, our classifier achieves high accuracy in classifying the bearing degradation stages.

RELATED STUDIES

Existing research on bearing fault prediction has largely focused on predicting the remaining useful life (RUL). However, in PHM for bearings, we are more concerned with accurately determining a bearing's current life stage at a given time. Therefore, this section discusses research related to the segmentation of bearing degradation stages.

In the PHM domain of bearings, accurate segmentation of the bearing degradation stages is crucial for achieving precision maintenance and management. Unlike solely predicting the RUL, accurately segmenting the degradation stages of bearings enables appropriate maintenance strategies and actions to be identified, such as implementing preventive maintenance or corrective measures at different stages. Researchers have proposed various methods and techniques to precisely segment bearing degradation stages. Among them, approaches based on image segmentation, deep learning, and unsupervised learning have played a significant role in research on bearing degradation stage segmentation.

Methods based on image segmentation transform vibration data into an image format and utilise image segmentation algorithms, such as threshold segmentation and edge detection, to separate degraded regions from normal regions in the vibration images (Liao et al., 2022; Wang et al., 2017). However, this approach can only achieve segmentation for two stages and may not accurately differentiate complex bearing degradation processes, such as the early stages of degradation. Deep learning-based methods using convolutional neural networks (CNNs) (Wang et al., 2021; Zhou et al., 2019) and RULNet (Gamanayake et al., 2023) have been used for feature learning and segmentation on vibration signals. These approaches can automatically learn feature representations and complex relationships, enabling precise bearing degradation segmentation. Nonetheless, deep learning methods typically require a large amount of labelled data for supervised learning, which poses challenges in engineering applications. Manually annotating a large-scale vibration dataset is time-consuming and labour-intensive, demanding domain expertise and experience correctly labelling samples of different degradation stages.

Segmentation methods based on unsupervised learning enable the segmentation of bearing degradation stages without requiring labelled data. For instance, methods such as generative adversarial networks (Mylonas & Chatzi, 2020) and variational auto-encoders (Hong et al., 2014) can learn the latent representation and distribution of data, enabling the segmentation of bearing degradation and dividing degraded areas into different categories. In unsupervised learning approaches, the key lies in modelling bearing degradation. Based on the nature of the degradation process, bearing degradation modelling methods can be categorised into continuous degradation and discrete degradation stage models.

The continuous degradation model aims to establish a single model to describe the gradual deterioration of the bearing's life. This model type is typically based on physical principles, statistics, or machine learning. The continuous degradation model considers the progressive changes in the bearing life and can capture the trends of bearing degradation based on various indicators (e.g. vibration features, temperature, and load). Common continuous degradation models include physical models (Cui & Su, 2021; Li et al., 2023), Markov processes models (Gu et al., 2023; Kou et al., 2022), and regression models (Chen et al., 2022). By estimating and predicting the parameters of a continuous degradation model, the model can forecast and assess the remaining life of a bearing. The establishment and prediction of the continuous degradation model are based on existing bearing degradation data and the assumption that the degradation process is continuous and gradual. However, the actual bearing degradation process may be influenced by multiple factors, such as working conditions, load variations, and environmental factors. These factors may result in discrepancies between the model and actual scenarios.

In the discrete degradation stage model, the bearing lifespan is divided into multiple discrete stages, based on which the bearing conditions are modelled and predicted. This model can adapt better to the nonlinearity and uncertainty of degradation processes in practical scenarios. Common discrete degradation stage models include hidden Markov models (Aggab et al., 2022) and time-series models (Cao et al., 2021; Zhu et al., 2023). These models can better describe and identify the different stages in the bearing degradation process, thereby assisting in determining appropriate maintenance strategies and operational measures.

The bearing discrete degradation stage model proposed by Juodelyte et al. (2022) transforms the life prediction into a classification problem. However, this model fails to identify effective features, resulting in lower accuracy in predicting early-stage bearing degradation. In contrast, our proposed approach introduces a convolutional autoencoder. It combines it with a domain-knowledge-based automatic data labelling strategy and a classifier based on a multi-scale attention network architecture. This model is suitable for predicting degradation throughout a bearing's lifespan.

METHODS

Bearing degradation can be categorised into four stages based on the observed physical manifestations and features in the frequency and time domains. This classification helps us better understand the changing trends and status evolution of bearing life. Each stage exhibits different degradation characteristics and requires specific maintenance strategies to prolong the lifespan of the bearing and ensure reliable equipment operation. Table 1 summarises the characteristics and corresponding maintenance strategies for each stage.

Table 1
Bearing degradation stage classification

Stage	Physical Characteristics	Time-domain Characteristics	Frequency-domain Characteristics	Maintenance Strategy
Healthy	The bearing is in normal operating condition.	It is manifested as stable periodic oscillations.	Normal operating frequency components are displayed in the spectrum.	Monitor and record the baseline condition and performance parameters of bearings.
Early degeneration	Minor signs of degradation, e.g. scratches, tiny cracks, or slight wear.	Minor non-periodic variations may occur, e.g. slight changes in amplitude.	High-frequency components related to early-stage faults appear in the spectrum.	1. Monitor spectrum data and observe early signs of fault variations. 2. Perform regular vibration analysis and diagnosis. 3. Timely lubricate and clean.
Severe degeneration	More significant and severe degradation.	Manifests as evident non-periodic fluctuations, potentially accompanied by more amplitude variations and waveform distortion.	The spectrum has more high-frequency components and wide-band noise associated with bearing faults.	1. Strengthen vibration analysis and fault diagnosis. 2. Implement appropriate maintenance and replacement measures. 3. Inspect lubrication conditions and the performance of lubricants.

Table 1 (*continue*)

Stage	Physical Characteristics	Time-domain Characteristics	Frequency-domain Characteristics	Maintenance Strategy
Failure	Severe fault state	Significant mutations, shocks, or irregular fluctuations may occur.	More high-frequency components and wide-band noise may appear in the spectrum, possibly causing impact frequencies.	1. Inspection and replacement of bearings and their related components. 2. Root cause analysis to prevent similar failures from occurring again.

This paper collects time-domain vibration data during the operation of bearings for bearing life-stage prediction. The problem is divided into two stages. Assuming the input is $X \in Rk, t = (0,1,\cdots,T)$, where T is the time step. X contains the complete life cycle data of the bearings. The data transformed into the frequency domain is represented as $F(X) \in Rm$, where m is the dimension of frequency domain features. The model maps the frequency domain features to bearing life-stage labels $L \in \{0,1,2,3\}$ to form life labels, reflecting different degradation states of the bearings. Equation 1 represents Stage I:

$$L = M(F(X)) \quad [1]$$

M represents the model function for Stage I.

Assuming the input data are $X \in Rk$, and $L \in \{0,1,2,3\}$, representing different life-stage labels of the bearings, the model takes input data X, which include the frequency and time domain features of the bearings, along with life-stage labels L. The output is the predicted life-stage label of the bearings, denoted as $P \in \{0,1,2,3\}$. Equation 2 represents Stage II:

$$P = G(X,L) \quad [2]$$

G represents the classifier for Stage II.

The functions M in Stage I and G in Stage II can be used to classify and predict the degradation stage of the bearing vibration data. This aids in determining the bearing life stage at a given time point and provides corresponding maintenance strategies. Figure 1 shows the workflow of our model, which is divided into two parts based on the description of the aforementioned problem. In the following, we discuss each submodel in detail.

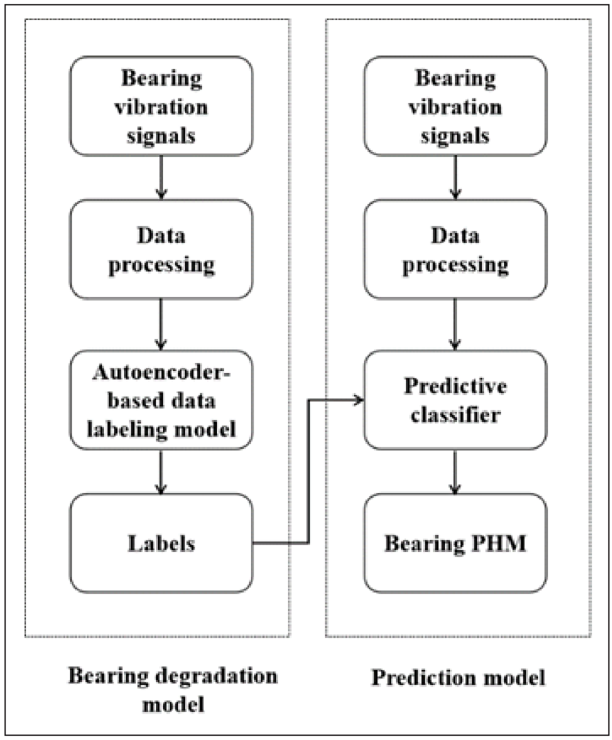

Figure 1. Bearing life-stage workflow

Bearing Degradation Model

The proposed bearing degradation model, FAEK, consists of data pre-processing, a convolutional autoencoder, and a clustering operation. The architecture of the FAEK network is shown in Figure 2.

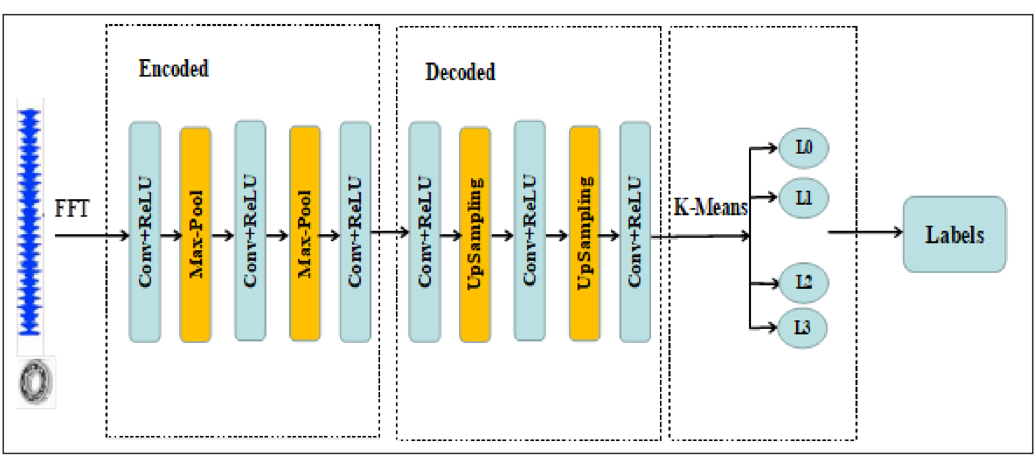

Figure 2. FAEK network architecture

First, the bearing vibration data are subjected to a fast Fourier transform (FFT) to determine the vibration signal's dominant frequency components and energy distribution, thereby extracting features related to bearing degradation. These features can be used to divide the degradation stages. Assuming that the bearing vibration data are represented as a time series $x(t)$, where t represents the time, the frequency-domain transformation can be expressed as Equation 3:

$$X(f) = ABS(FFT(x(t))) \quad [3]$$

$X(f)$ is the frequency-domain representation of the vibration data, f denotes the frequency, and ABS is the absolute value.

The frequency-domain signals of the bearings are fed into the convolutional autoencoder to learn the feature representation of the data through unsupervised learning. The convolutional autoencoder consists of an encoder and a decoder. The encoder compresses the input data into a latent representation, and the decoder decodes the latent representation into the reconstructed input data. In contrast to traditional fully connected autoencoders, convolutional autoencoders utilise convolutional operations in the encoding and decoding processes. In the bearing degradation model, the encoder uses 32, 64, and 128 convolutional kernels to sample the pattern information thrice. Then, it samples 64, 32, and 1 convolutional kernel in the decoder to recover the data thrice. The convolution kernel size for both the encoder and decoder is 3.

The encoder takes the input data and processes them through convolution and activation operations. It gradually reduces the size of the feature maps through max-pooling operations and extracts and compresses feature information from the input data. The results of the convolutional and pooling layers are denoted as h and p, respectively, in Equations 4 and 5:

$$h = ReLU(Conv1D(W, b)(x)) \quad [4]$$

$$p = MaxPooling1D(h) \quad [5]$$

$Conv1D$ represents a one-dimensional convolution operation; W and b are the weights and bias parameters corresponding to the convolutional layer; $ReLU$ represents the activation function; $MaxPooling$ represents a one-dimensional max-pooling operation.

The decoder gradually restores the size of the feature maps through upsampling operations, followed by convolution and activation operations, to obtain the reconstructed output. $UpSampling$ can be represented as u in Equation 6:

$$u = UpSampling1D(h) \quad [6]$$

$UpSampling1D$ represents the one-dimensional up-sampling operation.

Using a convolutional autoencoder, we can capture meaningful features from bearing vibration data and transform them into a meaningful representation. The convolutional autoencoder is trained in an unsupervised learning manner, eliminating the necessity for manual labelling of the bearing data's life stages and avoiding reliance on handcrafted feature engineering. The design of the convolutional autoencoder enables it to perform exceptionally well in processing the time-series data of bearing vibrations, enabling it to extract temporal correlations and local features from the data.

Finally, by mapping the learned features into a clustering space, the bearing data can be effectively grouped into different degradation stages, generating the corresponding labels. It provides essential guidance for subsequent maintenance decisions. This data-driven approach exhibits a good generalisation performance and can adapt to various types and scales of bearing data, making it a viable solution for practical engineering applications.

Prediction Model

Early signs of bearing degradation are only visible in the frequency domain (Qiu et al., 2023). We propose a multi-input classification network architecture, which we refer to as the multi-scale channel-attention classification (MCC) model, to capture both the frequency-domain features associated with early degradation signs and time-domain features that intensify with the progression of degradation. Figure 3 shows a schematic of this architecture.

In this architecture, we simultaneously input the time domain, frequency domain, and degradation signals. Time-domain signals provide more detailed temporal information, whereas frequency-domain signals can better capture the frequency characteristics. By combining these two inputs, we can fully exploit their complementarities. The network has three layers: multi-scale processing, channel attention, and classification. The substructures are discussed in detail in the following.

The multi-scale processing layer consists of a multi-branch parallel convolution, utilising four parallel convolutional branches to extract different frequency domain features. Each branch extracts a set of feature maps, capturing distinctive aspects of the input data. These feature maps are then fused to obtain a more comprehensive representation of the features. This approach provides a richer and more diverse feature representation, enabling the network to adapt to various bearing degradation patterns and enhance its ability to extract complex features.

If the input data are denoted as X, the number of convolution kernels is N, the size of each convolution kernel is $K \times K$, the number of input data channels is C, and the number of output feature maps is M. The weight matrix of each convolution kernel $I\,(1 \leq i \leq N)$ is represented as W_n with the shape $K \times K \times C$, and its bias term is b_n. The output of the multi-path convolution operation for the m-th feature map of the input data X can be calculated using Equation 7:

$$Y_m = \sum^{n} f(\mathrm{conv}(W_n, X) + b_n) \qquad [7]$$

Here, $conv(W_n, X)$ denotes the convolution operation of input data X with convolution kernel W_n, b_n denotes the bias term of the n-th convolution kernel, and f denotes the ReLU function.

During the bearing degradation process, the frequency spectrum characteristics of the vibration signal change. Traditional single convolutional kernels often struggle to capture subtle variations across frequency bands. By contrast, multiple parallel convolutions, known as multi-branch parallel convolutions, employ several parallel convolutional kernels, each corresponding to a different frequency band. It enables better extraction of signal features in the frequency domain. Moreover, multi-branch parallel convolution increases the depth and width of the network, providing more parameters and nonlinear expressive power. Consequently, the model can better adapt to complex vibration signal features, improving its bearing degradation research performance.

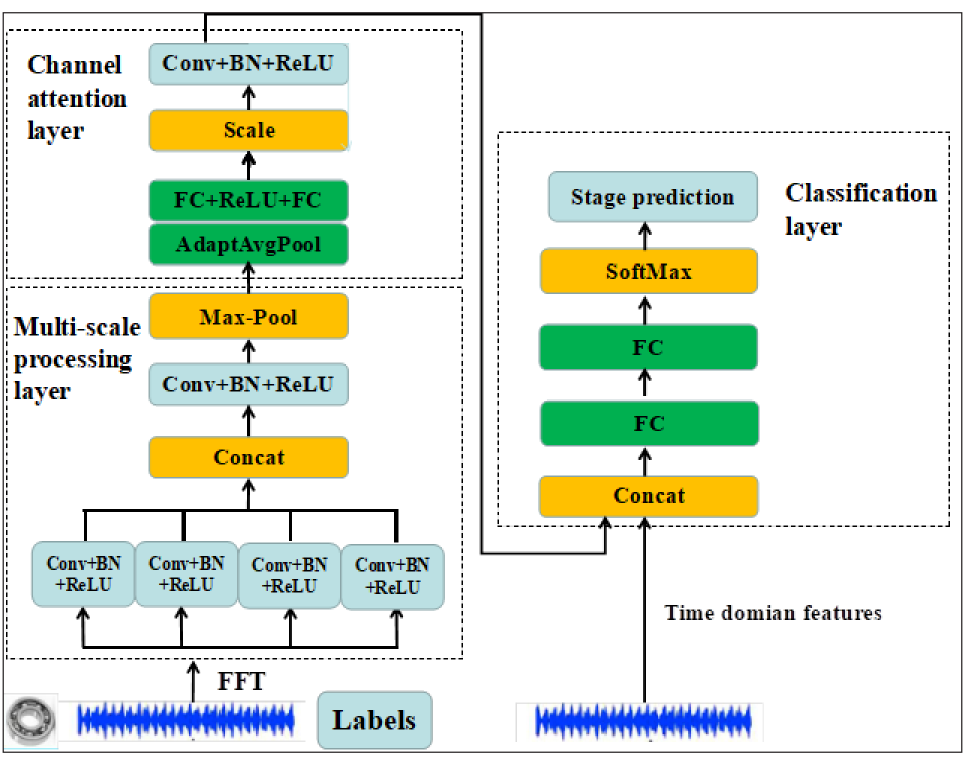

Figure 3. MCC architecture

Channel Attention Layer

The channel attention network won the championship in the ImageNet classification task in 2017 (Hu et al., 2019). It can enhance the attention mechanism of CNNs by adapting the importance of features across channels to improve network performance.

The channel attention module comprises two key steps: squeezing and excitation. In the squeezing step, the input feature map undergoes global pooling along the channel dimensions to obtain global information regarding each channel. In the excitation step, a nonlinear function is learned to transform the squeezed features into channel attention weights that adjust each channel's importance in the input feature map. Figure 4 shows the structure of the channel attention module.

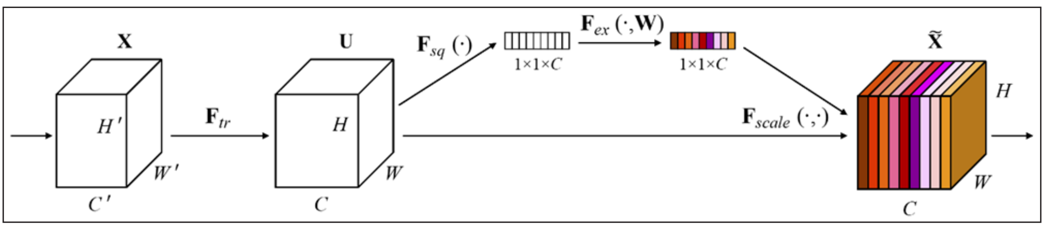

Figure 4. Channel attention module (Hu et al., 2019)

Squeezing: If the input feature map has dimensions H, W, and C, the compression operation applies global average pooling to the input feature map, resulting in a C-dimensional vector that represents the global average value for each channel. This vector indicates the overall importance of each channel in the input-feature map. The output result Z can be represented as Equation 8:

$$Z = F_{sq}(X) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X(i,j) \quad [8]$$

Excitation: The excitation operation generates channel attention weights by learning a nonlinear mapping function. The mapping function can be learned using a fully connected or one-dimensional convolutional layer. Given an input feature vector Z, the result of the excitation operation, denoted as S, can be represented as Equation 9:

$$S = \sigma(W2 * \delta(W1 * Z)) \quad [9]$$

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

Scale: The channels of X are reweighted by performing element-wise multiplication between S and X. Thus, the weighted feature map Y is calculated as Equation 10:

$$Y = \mathrm{Fscale}(S, X) = S * X \tag{10}$$

By learning the weights of each channel, the channel attention network can adaptively enhance important features. The vibration frequency-domain signals of the bearings can be used to extract the most representative and discriminative frequency-domain features, which is beneficial for accurately analysing the degradation of bearings. Moreover, the channel attention network can significantly improve the network performance without introducing additional parameters. It is particularly important for large-scale data processing in bearing degradation research because bearing vibration signals are typically high-dimensional time-series data. The network can efficiently and accurately extract features without increasing the network complexity.

Classification Layer

In bearing fault diagnosis, the time-domain features are crucial for capturing fault conditions. Skewness and kurtosis are commonly used time-domain features that reflect vibration signals' symmetry and impulsive characteristics. These features are essential for predicting the degradation of bearings. When the bearings are faulty, the pulse components in the vibration signal typically increase, making these features valuable for bearing fault diagnosis. Additionally, dimensionless parameters can mitigate the influence of operating conditions and other factors on the model. Common dimensionless parameters include the clearance, waveform, impulse, peak, and kurtosis indices. The clearance index is often used to indicate wear conditions. In contrast, the peak, impulsion, and kurtosis indices are used to describe the impact characteristics of the signal, aiding in better characterising early degradation. Table 2 presents the 13 time-domain features that we extracted (Zhu et al., 2014).

To identify bearing degradation more accurately, the feature dimensions of the vibration signal were reduced after performing feature extraction with the aforementioned two-layer structure. By combining frequency-domain and time-domain features and then applying linear layers and softmax for degradation-stage prediction, this fusion approach enables comprehensive utilisation of information from both feature categories, thereby enhancing diagnostic and predictive accuracy. Thus, to predict the bearing stage model, the convolution operation of the input layer uses 32 convolution kernels for sampling, all with a size of 3 and an activation function of ReLU.

Table 2
Time-domain characteristics (Zhu et al., 2014)

Sequence	Parameter	Sequence	Parameter	Sequence	Parameter
1	Peak $x_p = max\lvert x_i\rvert$	2	Cliffness $K = \frac{\beta}{x^4_{rms}}$	3	Average square root $x_{rms} = \sqrt{1/N\sum_{i=1}^{N} x_i^2}$
4	Average $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$	5	Yield indicator $L = \frac{x_p}{x_r}$	6	Fang root amplitude $x_r = (1/N\sum_{i=1}^{N} \sqrt{\lvert x_i\rvert})^2$
7	Average amplitude $\lvert\overline{x}\rvert = \frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert$	8	Waveform indicators $S = \frac{x_{rms}}{\lvert\overline{x}\rvert}$	9	Slope $\alpha = 1/N\sum_{i=1}^{N} x_i^3$
10	Variance $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \overline{x})^2$	11	Pulse indicator $I = \frac{x_p}{\overline{x}}$	12	Cliffness indicator $\beta = 1/N\sum_{i=1}^{N} x_i^4$
13	Peak indicators $C = \frac{x_p}{x_{rms}}$				

RESULTS AND DISCUSSION

We used a publicly available bearing dataset to validate the proposed model's performance. We evaluated automatic label assignment and compared the classification results for different bearings under various operating conditions. Additionally, we assessed the accuracy of predicting the bearings' degradation stages and made comparisons to demonstrate the effectiveness of our model in predicting bearing degradation. We aimed to understand the effectiveness of the model in generalising across different types of bearings and operating conditions.

Datasets Description

This study employed the FEMTO-ST bearing dataset for training and evaluation (Zhang et al., 2018). This dataset contains horizontal and vertical vibration measurement data of ball bearings with rolling elements collected through experiments. The bearings were tested under pressure exceeding the recommended load to accelerate the degradation process. As shown in Table 3, 17 bearings were tested under three different load and speed conditions. For safety reasons, the experiment stops when the accelerometer reading exceeds 20g, indicating that the bearing has reached the final degradation stage. Due to the high variability of bearing life (from 28 minutes to 7 hours), the FEMTO bearing dataset is challenging (Figure 5). There is no specific explanation for the bearing malfunction. Even worse, the data challenge description mentions that bearings may have multiple defects occurring simultaneously. During the experiment, the vibration data of the bearing was sampled every 10 seconds at a frequency of 25600Hz for 0.1 seconds. As a result, each sample contained 2560 × 2 data points.

Table 3
FEMTO-ST dataset (Zhang et al., 2018)

Working conditions	Load force (N)	Speed (r/min)	No. of training sets	No. of test sets
1	4000	1800	2	5
2	4200	1650	2	5
3	5000	1500	2	1

Data Preparation and Experimentation

First, we pre-processed raw data from the FEMTO-ST dataset. We independently separated and processed the data for each direction because the dataset contains vibration signals for horizontal and vertical directions. We applied an FFT to convert the vibration signal in each direction into the frequency domain and extract the frequency-domain features. Additionally, we extracted the time-domain features for each direction's vibration signal based on the feature list provided in Table 2, described their statistical and impulsive characteristics, and then fused the features from horizontal and vertical directions to obtain comprehensive and integrated bearing features. Because each bearing may have different operating conditions and degradation patterns, we performed all the above operations separately to capture its unique vibration characteristics and degradation patterns.

We used cross-entropy as the loss function and the Adam optimiser to train the model. During the iterations, we employed gradient clipping techniques to prevent gradient explosions. The batch size for all the networks was 64, and the learning rate was 0.01. All the experiments were conducted on a Windows 10 operating system with an Intel i5-12490F CPU and an NVIDIA RTX 3060 12 GB GPU. Our code was written in Python 3.8 using PyTorch.

To evaluate the effectiveness of the automatic label assignment of the model, we compared automatically generated stage labels with manually assigned stages from for the training set (Sutrisno et al., 2012). Smaller differences indicated better accuracy in the model label assignment. Additionally, to assess the performance of the bearing stage prediction classifier, we compared the predicted stages with the stages automatically labelled by our model for the test set to evaluate the prediction accuracy. As bearing PHM requires predictions to follow the actual degradation sequence over time, we also considered the overlap between predicted and actual stages when predictions were made in chronological order, which aided in gauging the model's reliability.

Experimental Results

Data labelling: We automatically assigned degradation stages to the six bearings in the training set and compared the results with those of the manually labelled stages. As shown in Figure 5, the model's automatically assigned bearing degradation stages were in good agreement with the manually labelled stages, indicating that the model performed well in identifying the bearing degradation stages. Although some overlap was observed between adjacent stages, this was reasonable because of the continuity of the degradation process and variations in the signals, which may have resulted in some degree of overlap between neighbouring degradation stages. The model demonstrated excellent performance in distinguishing between healthy and faulty stages, which is crucial for early maintenance and timely identification of failed bearings. The FEMTO-ST dataset was collected under extreme operating conditions, resulting in relatively short healthy stages for all bearings, which aligned with real-world scenarios.

However, for bearing1_2, the model's performance was relatively poor. This was owing to the excessive amplitude of the smoothest maximum acceleration in the vertical direction, which resulted in strong interference in the signal. Despite this, the model exhibited good accuracy and robustness in automatically assigning bearing degradation stages.

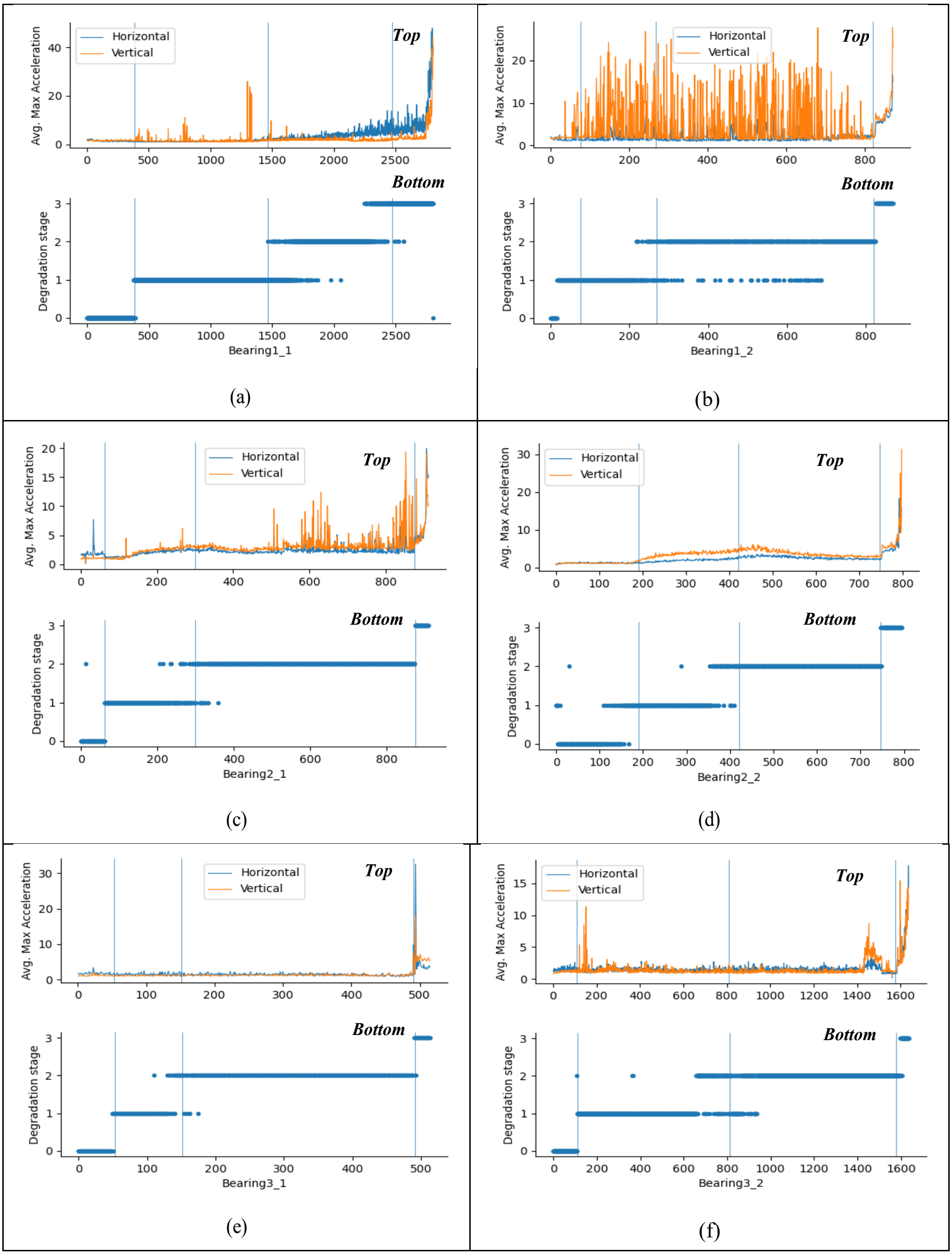

Figure 5. The training set degradation labels: Top: Smoothed maximum acceleration obtained by averaging the five highest absolute acceleration measurements in the time domain (Sutrisno et al., 2012). Bottom: Automatically assigned labels obtained using the FAEK model. The vertical lines represent the manually assigned results

Classifier

After automatically labelling the training set for the six bearings using the FAEK model, we compared the performances of the AE model (Juodelyte et al., 2022), 1DCNN model (Wang et al., 2021), and our proposed MCC model in predicting the bearing degradation stages (Figure 6).

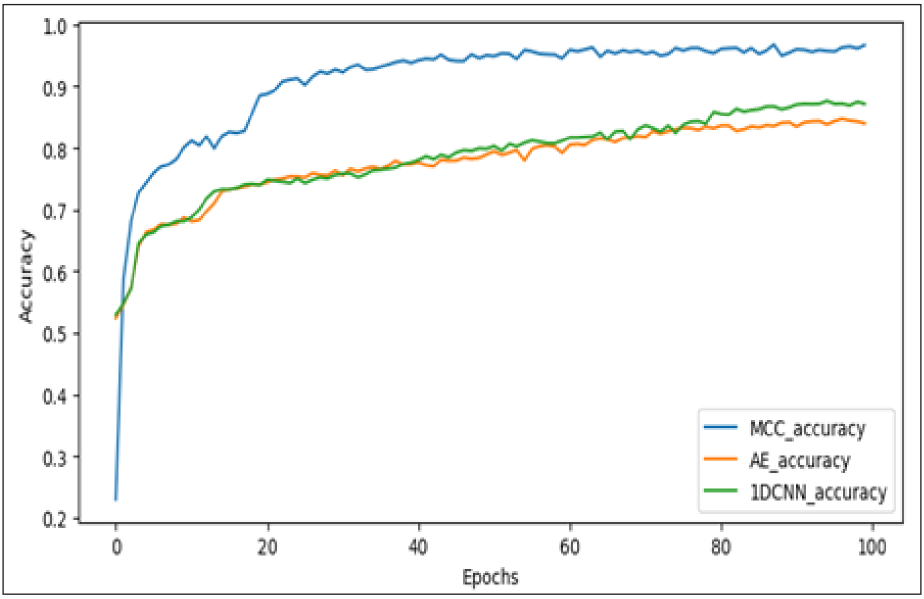

Figure 6. Accuracy of different classification models

The results indicated that the MCC model outperformed the other two with remarkable classification accuracy for the bearing degradation classification task, reaching approximately 0.9665. The superiority of the MCC model can be attributed to its innovative design and feature extraction techniques. The model adopts multi-path convolution and attention mechanisms to effectively extract meaningful features. Additionally, we simultaneously trained the model on both the horizontal and vertical data of the bearings and ultimately fused the frequency- and time-domain features.

The AE and 1DCNN models exhibited similar performances, with maximum accuracies of 0.8471 and 0.8763, respectively. Notably, our proposed MCC model achieved an accuracy greater than 0.9 after only 30 epochs, indicating its high convergence speed. This efficiency makes the MCC model more effective in solving the bearing degradation classification task, thereby providing timely and accurate solutions for bearing fault diagnosis and prediction.

To validate the accuracy of our model in predicting the bearing degradation stages further, we compared the results of the MCC and 1DCNN models on the test set (Figure

7). The left-hand figure shows the prediction accuracy of the 1DCNN model. The box plot reveals that the 1DCNN model had a lower classification accuracy in stages 1 and 3, with the median line of the boxes being relatively low and the boxes narrow, indicating a significant dispersion of data in these two stages. In particular, the performance was poor in bearing degradation stage 1. The right-hand figure in Figure 7 shows the results of our MCC model, indicating a high classification accuracy at all stages. The median line of the boxes was higher, and the boxes were flatter, suggesting a higher concentration of data at each stage. Therefore, the MCC model exhibited more stable performance and superior results.

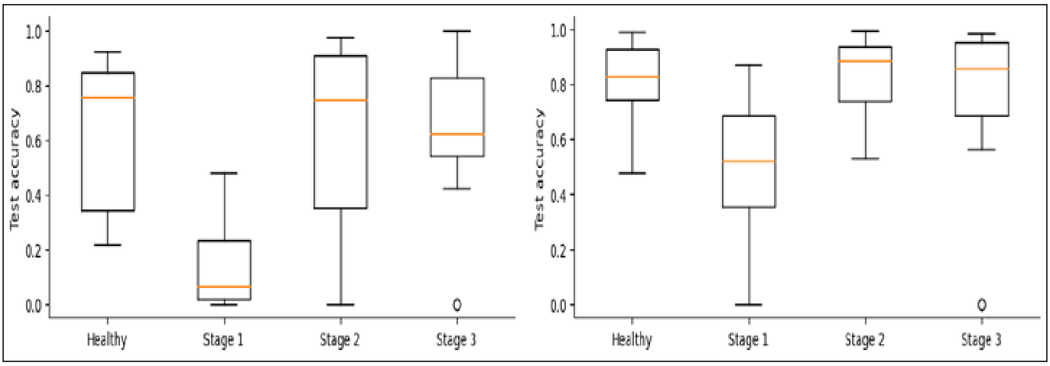

Figure 7. Test set accuracy of the models (based on FAEK labels). Left: 1DCNN model. Right: MCC model

Figure 8 shows the statistics on the overlapping rates between the adjacent stages of all 11 bearings in the test set. Our model's predicted results had a low overlapping rate, indicating the MCC model's ability to distinguish different degradation stages and better identify the distinctive features of each stage in the bearing degradation classification task.

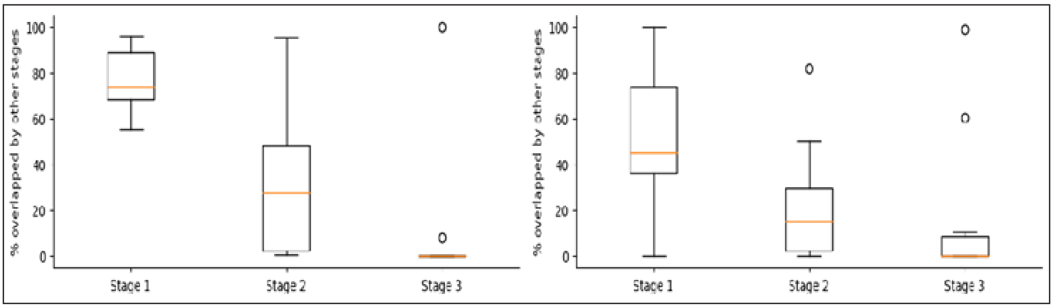

Figure 8. Overlap rate of adjacent stages for the models. Left: 1DCNN Model. Right: MCC Model

We presented the prediction results using a smoothing classifier to reduce prediction fluctuations at individual time points and obtain more stable predictions. Specifically, we generated the posterior probabilities for five predictions for each time point in the test set.

We calculated their average within a fixed sliding window, resulting in smooth classifier predictions.

Figure 9 shows the prediction results for three different operating conditions for the test set containing 11 bearings. The graphs show that the predicted bearing degradation stages aligned well with the actual bearing degradation progress over time, indicating that our predictions were highly accurate and consistent with the actual degradation processes of the bearings.

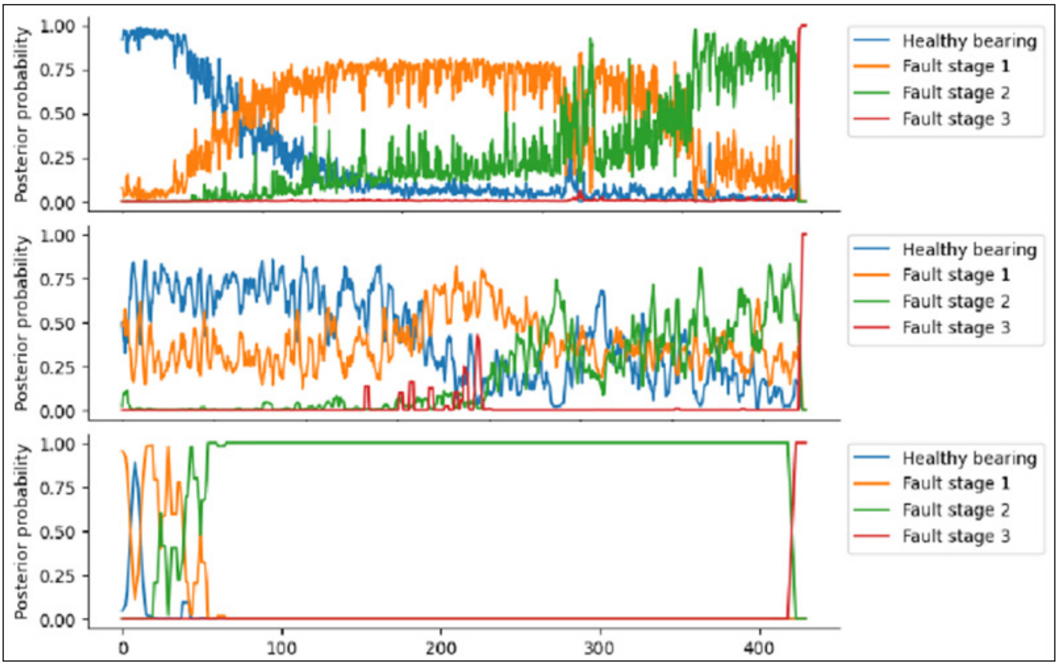

Figure 9. Prediction results for bearings 1_5 (top), 2_4 (middle), and 3_3 (bottom)

CONCLUSION

This study focused on the problem of predictive maintenance for bearings, aiming to classify bearings into four degradation stages to predict at a given time point when the bearing is in the degradation stage and to provide corresponding maintenance strategies for different stages. We propose a method based on a FAEK automatic label assignment and an MCC prediction model to achieve this. This model has the following characteristics: (1) automatic label assignment: We use a convolutional autoencoder-based approach to segment the full-lifetime vibration signals of bearings into different stages. The FAEK automatic label assignment effectively distinguishes different degradation stages of bearings, avoiding the laborious and subjective nature of manual label assignment, and (2) prediction model: We input both time- and frequency-domain features of bearing signals into the MCC classification model, which includes multiple convolutional modules and channel attention modules. These modules aid in better extracting signal features and achieving a precise prediction of the bearing degradation stages. With this prediction model, we can accurately predict the degradation status of bearings and promptly perform corresponding maintenance measures. We validated this network architecture using the FEMTO dataset. The results revealed excellent performance. The automatically assigned labels highly overlapped with the manually assigned labels, successfully distinguishing different degradation stages. For degradation-stage prediction, our model architecture achieved an accuracy of 0.9665.

The bearing dataset used in this study is relatively clean, whereas in practical industrial applications, bearing data may be subject to various types of interference and noise. Additionally, the accuracy of the automatically assigned labels requires further discussion and improvement. Future research should explore better segmentation methods to improve the quality of labels. Only by improving the quality of the labels can our classification prediction results become more trustworthy and reliable, thereby providing more effective support for the predictive maintenance of bearings.

ACKNOWLEDGEMENT

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

Aggab, T., Vrignat, P., Avila, M., & Kratz, F. (2022). Remaining useful life estimation based on the joint use of an observer and a hidden Markov model. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *236*(5), 676–695. https://doi.org/10.1177/1748006X211044343

Cao, Y., Jia, M., Ding, P., & Ding, Y. (2021). Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network. *Measurement*, *178*, Article 109287. https://doi.org/10.1016/j.measurement.2021.109287

Chen, Y., Duan, W., Ding, Z., & Li, Y. (2022). Battery life prediction based on a hybrid support vector regression model. *Frontiers in Energy Research*, *10*, Article 899804.

Cui, L., & Su, Y. (2021). Contact fatigue life prediction of rolling bearing considering machined surface integrity. *Industrial Lubrication and Tribology*, *74*(1), 73–80. https://doi.org/10.1108/ILT-08-2021-0345

Gamanayake, C., Qin, Y., Yuen, C., Jayasinghe, L., Tan, D.-E., & Low, J. (2023). A hybrid deep learning model-based remaining useful life estimation for reed relay with degradation pattern clustering. *IEEE Transactions on Industrial Informatics*, *19*(6), 7401–7413. https://doi.org/10.1109/TII.2022.3210250

Gu, M. Y., Ge, J. Q., & Li, Z. N. (2023). Improved similarity-based residual life prediction method based on grey Markov model. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, *45*(6), Article 294. https://doi.org/10.1007/s40430-023-04176-z

Hong, S., Zhou, Z., Zio, E., & Wang, W. (2014). An adaptive method for health trend prediction of rotating bearings. *Digital Signal Processing*, *35*, 117–123. https://doi.org/10.1016/j.dsp.2014.08.006

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2019). *Squeeze-and-Excitation Networks* (arXiv:1709.01507). arXiv. https://doi.org/10.48550/arXiv.1709.01507

Juodelyte, D., Cheplygina, V., Graversen, T., & Bonnet, P. (2022). Predicting bearings degradation stages for predictive maintenance in the pharmaceutical industry. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3107-3115). ACM Publishing. https://doi.org/10.1145/3534678.3539057

Kou, L., Chu, B., Chen, Y., & Qin, Y. (2022). An automatic partition time-varying markov model for reliability evaluation. *Applied Sciences*, *12*(12), Article 12. https://doi.org/10.3390/app12125933

Li, H. F., Wei, J. L., Li, S. H., Liu, Y. Q., Gu, X. H., Liu, Z. C., & Yang, S. P. (2023). Fatigue life prediction of high-speed train bearings based on the generalized linear cumulative damage theory. *Fatigue & Fracture of Engineering Materials & Structures*, *46*(6), 2112–2120. https://doi.org/10.1111/ffe.13984

Li, X., Zhang, W., & Ding, Q. (2019). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliability Engineering & System Safety*, *182*, 208–218. https://doi.org/10.1016/j.ress.2018.11.011

Liao, D., Yin, M., Luo, H., Li, J., & Wu, N. (2022). Machine vision system based on a coupled image segmentation algorithm for surface-defect detection of a Si_3N_4 bearing roller. *JOSA A*, *39*(4), 571–579. https://doi.org/10.1364/JOSAA.449088

Mylonas, C., & Chatzi, E. (2020). *Remaining Useful Life Estimation Under Uncertainty with Causal GraphNets* (arXiv:2011.11740; Version 1). arXiv. http://arxiv.org/abs/2011.11740

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), Article 3. https://doi.org/10.3390/s23031305

Ren, L., Sun, Y., Cui, J., & Zhang, L. (2018). Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. *Journal of Manufacturing Systems*, *48*, 71–77. https://doi.org/10.1016/j.jmsy.2018.04.008

Ruan, D., Han, J., Yan, J., & Gühmann, C. (2023). Light convolutional neural network by neural architecture search and model pruning for bearing fault diagnosis and remaining useful life prediction. *Scientific Reports*, *13*(1), Article 1. https://doi.org/10.1038/s41598-023-31532-9

Sanakkayala, D. C., Varadarajan, V., Kumar, N., Karan, Soni, G., Kamat, P., Kumar, S., Patil, S., & Kotecha, K. (2022). Explainable AI for bearing fault prognosis using deep learning techniques. *Micromachines*, *13*(9), Article 9. https://doi.org/10.3390/mi13091471

Sutrisno, E., Oh, H., Vasan, A. S. S., & Pecht, M. (2012). Estimation of remaining useful life of ball bearings using data driven methodologies. In *2012 IEEE Conference on Prognostics and Health Management* (pp. 1-7). IEEE Publishing. https://doi.org/10.1109/ICPHM.2012.6299548

Wang, B., Pan, H., & Yang, W. (2017). Robust bearing degradation assessment method based on improved CVA. *IET Science, Measurement & Technology, 11*(5), 637-645. https://doi.org/10.1049/iet-smt.2016.0391

Wang, C., Jiang, W., Yang, X., & Zhang, S. (2021). RUL prediction of rolling bearings based on a DCAE and CNN. *Applied Sciences*, *11*(23), Article 23. https://doi.org/10.3390/app112311516

Wang, F., Liu, X., Deng, G., Yu, X., Li, H., & Han, Q. (2019). Remaining life prediction method for rolling bearing based on the long short-term memory network. *Neural Processing Letters*, *50*(3), 2437–2454. https://doi.org/10.1007/s11063-019-10016-w

Wang, Y., Deng, L., Zheng, L., & Gao, R. X. (2021). Temporal convolutional network with soft thresholding and attention mechanism for machinery prognostics. *Journal of Manufacturing Systems*, *60*, 512–526. https://doi.org/10.1016/j.jmsy.2021.07.008

Xia, T., & Xi, L. (2019). Manufacturing paradigm-oriented PHM methodologies for cyber-physical systems. *Journal of Intelligent Manufacturing*, *30*(4), 1659–1672. https://doi.org/10.1007/s10845-017-1342-2

Zhang, C., Pan, L., Wang, S., Wang, X., & Tomovic, M. (2018). An accelerated life test model for solid lubricated bearings used in space based on time-varying dependence analysis of different failure modes. *Acta Astronautica*, *152*, 352–359. https://doi.org/10.1016/j.actaastro.2018.08.027

Zhou, Q., Shen, H., Zhao, J., Liu, X., & Xiong, X. (2019). Degradation state recognition of rolling bearing based on K-means and CNN algorithm. *Shock and Vibration*, *2019*, Article e8471732. https://doi.org/10.1155/2019/8471732

Zhu, G., Zhu, Z., Xiang, L., Hu, A., & Xu, Y. (2023). Prediction of bearing remaining useful life based on DACN-ConvLSTM model. *Measurement*, *211*, Article 112600. https://doi.org/10.1016/j.measurement.2023.112600

Zhu, J., Nostrand, T., Spiegel, C., & Morton, B. (2014). Survey of condition indicators for condition monitoring systems. *Annual Conference of the PHM Society*, *6*(1), Article 1. https://doi.org/10.36001/phmconf.2014.v6i1.2514