

Bearing fault diagnosis method based on Gramian angular field and ensemble deep learning

Yanfang Han¹, Baozhu Li², Yingkun Huang³, Liang Li⁴

¹Sichuan College of Architectural Technology, Chengdu, 610399, China

²Internet of Things and Smart City Innovation Platform, Zhuhai Fudan Innovation Institute, Zhuhai, 518057, China

³High Performance Computing Department, National Supercomputing Center in Shenzhen, Shenzhen, 518055, China

⁴College of Electrical Engineering, Southwest Jiaotong University, Chengdu, 610031, China

²Corresponding author

E-mail: ¹hanyanfang839@163.com, ²tiger_1984@yeah.net, ³hykun@live.com, ⁴lianli@my.swjtu.edu.cn

Received 3 July 2022; received in revised form 24 August 2022; accepted 5 September 2022

DOI <https://doi.org/10.21595/jve.2022.22796>



Copyright © 2022 Yanfang Han, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract. Inspired by the successful experience of convolutional neural networks (CNN) in image classification, encoding vibration signals to images and then using deep learning for image analysis to obtain better performance in bearing fault diagnosis has become a highly promising approach. Based on this, we propose a novel approach to identify bearing faults in this study, which includes image-interpreted signals and integrating machine learning. In our method, each vibration signal is first encoded into two Gramian angular fields (GAF) matrices. Next, the encoded results are used to train a CNN to obtain the initial decision results. Finally, we introduce the random forest regression method to learn the distribution of the initial decision results to make the final decisions for bearing faults. To verify the effectiveness of the proposed method, we designed two case analyses using Case Western Reserve University (CWRU) bearing data. One is to verify the effectiveness of mapping the vibration signal to the GAFs, and the other is to demonstrate that integrated deep learning can improve the performance of bearing fault detection. The experimental results show that our method can effectively identify different faults and significantly outperform the comparative approach.

Keywords: bearing fault diagnosis, Gramian angular field, deep learning, ensemble learning.

1. Introduction

In modern industries, machine health monitoring is a prerequisite for maintaining the proper operation of industrial machines. Breakdowns in industrial machines can cause huge financial losses and even pose a threat to the people who use them. Therefore, the need for better and smarter machine health-monitoring technologies has never ceased [1]. Rolling bearings are considered the most common and critical mechanical components in rotating machinery, and their health can have a significant impact on the performance, stability, and service life of the machine. Because rolling bearings are usually in harsh operating environments, they are prone to failure during operation. Failure to detect defects in time can lead to unplanned machine downtime or even catastrophic damage. Therefore, rolling bearing fault detection is essential for the safe and reliable operation of machinery and production [2].

Recently, several bearing fault recognition methods have been proposed. Learning-based (including statistical learning methods and neural network methods) recognition methods can capture mechanical fault information by learning historical data and thereby enabling the automated analysis of bearing faults. The flow chart of these methods usually includes data preprocessing, feature extraction, and classifier design. Although a well-designed classification algorithm is a prerequisite for automated bearing fault detection [3], data preprocessing and feature extraction are also important steps.

Designing manual features based on the signal mechanism is a hot field for bearing fault

diagnosis. Chen et al. [4] merged the bearing signal features of the time and frequency domains and then inputted these features into a deep fully connected network for fault detection. Bao et al. [5] calculated the L-kurtosis feature in the envelope spectrum of a vibration signal to detect pulse periodicity. Chen et al. [6] first transformed a vibration signal into the spectrum domain and extracted the mapping amplitude entropy as a learnable feature. Zhao et al. [7] first performed empirical mode decomposition (EMD) on a vibration signal, and then selected the top few-mode components containing the main information of the signal to extract the sample entropy. Liu et al. [8] proposed a feature extraction method based on variational mode decomposition and singular value decomposition (SVD). Unfortunately, these manual feature extraction processes are laborious and unfriendly to machine-learning designers.

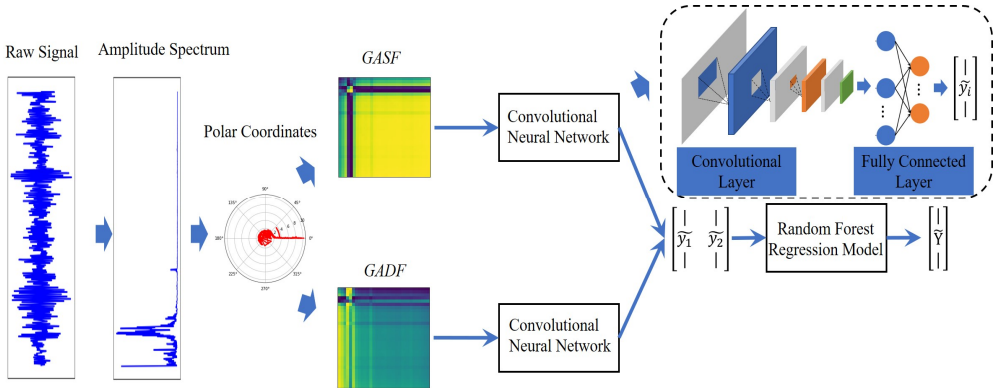


Fig. 1. The flow chart of the proposed method

Owing to the powerful feature learning ability of deep learning techniques, many researchers have attempted to introduce deep learning into the field of fault detection. Example include convolutional neural network (CNN)-based methods [1], [2], [9], [10]-[14], sparse auto-encoder (SAE)-based methods [15], [16], and recursive neural network (RNN)-based methods [17], [18], etc. Because the original signal is easily affected by noise, the signal is often transformed into an amplitude-frequency domain sequence. Generally speaking, CNN models are good at learning deep features from image data; thus, one-dimensional vibration signals encoded as two-dimensional image data have attracted much attention. CNN and their improved models have been successfully applied to image classification because they can extract robust features directly from two-dimensional images. Many image-interpreting vibration signal approaches have been proposed. Ding et al. [2] proposed a method for reconstructing a two-dimensional wavelet packet energy image (WPI) of the frequency space. The WPI can represent the dynamic structure of the wavelet packet energy distribution of different bearing faults. However, the WPI method combines the wavelet packet transform and phase space reconstruction technique, which not only has high time complexity but also loses the information of the original signal when performing multiple transformations of the representation space. Mantas et al. [19] converted the time series to the Permutation Entropy (PE) pattern, which is a 2D digital image with multiscale time delays. However, the method ignores the amplitude information of the time series, and it is sensitive to noise in the view of the principle behind the method. Wang et al. [20] combined Symmetrized Dot Pattern (SDP) with CNN for intelligent bearing fault diagnosis. SDP method [21] converts the time series into polar SDP images, which have the frequency and amplitude of the raw signals. In [22] and [9], they convert the time series into a 2D gray image. In addition, a variety of methods use the time-frequency image to represent the time series [23], [24], but the time-frequency analysis methods cost a lot of time so it is difficult for actual online diagnosis. Wang et al. [25] encoded a time series as the GAF image and then took advantage of the deep model for image representation learning to obtain a better classification accuracy rate. The algorithm was validated

on a 20-time series dataset and outperformed the traditional k-NN+DTW method, which was earlier considered the most effective method for time series classification.

The advantages of image-interpreting time series as a GAF image include the following: 1) there is no spatial transformation of the original time series, and the encoding process is performed in the original representation space. In other words, the time complexity of the algorithm is of the linear order $O(n)$, where n is the length of the time series. 2) The principle of the approach is simple, easy to understand, and reproducible. Surprisingly, the GAF pictorial method has rarely been applied to bearing fault detection problems. The GAF pictorial method has two different representations, including the Gramian Angular Summation Field (GASF) and the Differential Difference Field (GADF). The GASF and GADF provide different levels of information, such that, the final decision results of using them for deep learning are different. Given this, this study introduces the stacking generalization method [26], [27] to fuse the initial decision results based on GASF and GADF. The proposed method improved the classification accuracy rate of bearing faults and increased the reliability of the results.

The flow of the proposed approach is shown in Fig. 1. It can be seen that the method is composed of data preprocessing, image-interpreted vibration signal, and ensemble deep learning. The image-interpreting stage was used to better explore the vibration signals information. Ensemble deep learning can make full use of different levels of information to achieve better performance in bearing fault detection.

In sum, the main contributions of this study are as follows:

1) The GAFs approach is introduced to encode the bearing vibration signal into GASF and GADF matrices. The GAFs contain temporal correlation of vibration signal. The 2D-CNN is used to learn the deep features of the images, and then the idea of the stacking ensemble method is combined to construct an integrated deep model to achieve a high accuracy rate of bearing fault detection. It is a decision-level fusion strategy. Due to the deep learning with GASF and GADF obtains different accuracy rates, namely, they contain different information for classification. Therefore, building an ensemble model is an ideal fault detection scheme.

2) Last but not least, we design two experiments on CWRU datasets to evaluate the performance of bearing fault classification. With those comparable results, we demonstrate that our method achieves a better performance than the comparative method.

3) The rest of the paper is organized as follows: In Section 2, we introduce the principle of encoding time series as GAF images; we present the proposed method in Section 3; next, we conduct a performance test on the CWRU dataset. Finally, in Section 5, we conclude the paper.

2. Image-interpreted time series

Encoding vibration signals into images of different granularities is a popular research area for bearing fault diagnosis. Here, we introduce the GAF encoding method, which first transforms the vibrating bearing signal into a 2D image, and then uses a 2D-CNN to learn the knowledge of the image.

2.1. GAF encoding method

Given a time series $X = \{x_1, x_2, \dots, x_n\}$, where x_i is scaled in $[-1,1]$ or $[0,1]$ using:

$$\tilde{x}_i = \frac{x_i - \max(X) + x_i - \min(X)}{\max(X) - \min(X)}, \quad (1)$$

$$\tilde{x}_i = \frac{x_i - \min(X)}{\max(X) - \min(X)}. \quad (2)$$

Then, the \tilde{X} can be encoded as the angular cosine and the timestamp as the radius using Eq. (3):

$$\begin{cases} \phi = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{\mathbf{X}}, \\ r = \frac{t_i}{N}, & t_i \in \mathbb{N}, \end{cases} \quad (3)$$

where t_i is the timestamp and N is a constant factor. Based on this, a one-dimensional time series is mapped to the two-dimensional image. Because $x_i \in [-1,1]$ or $x_i \in [0,1]$, $\cos(\phi)$ ($\phi \in [0, \pi]$) is monotonic, such that GAF encoding is bijective. In other words, a time series can be mapped only to a unique polar coordinate space. In addition, the time dependence is preserved by the r coordinates.

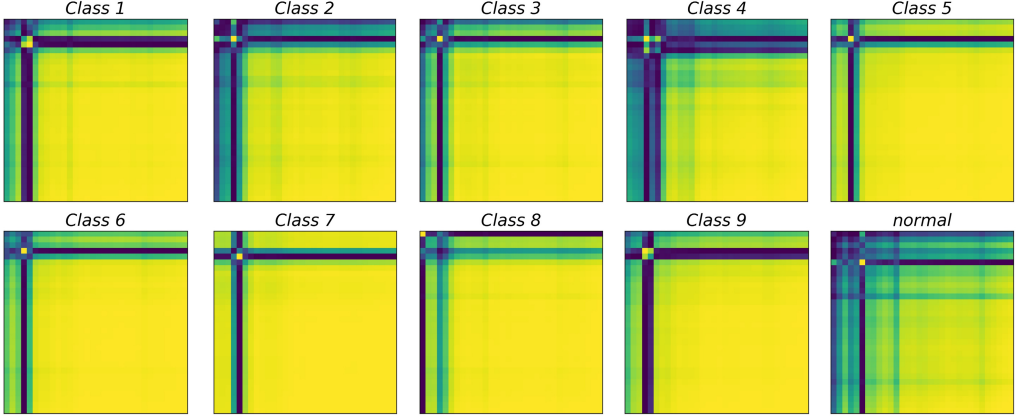


Fig. 2. GASF images of 10 bearing status

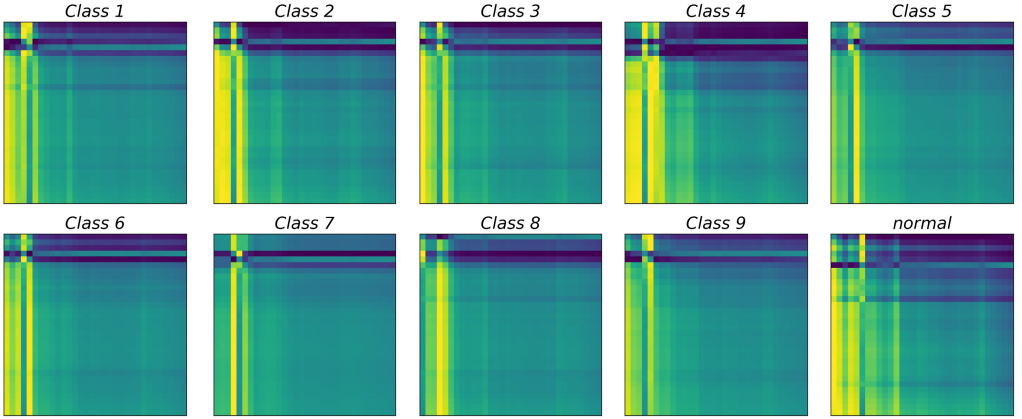


Fig. 3. GADF images of 10 bearing status

Once performing the GAF encoding for time series, the temporal correlations within different time intervals are identified by considering the triangular sum/difference between each point:

$$GASF(i, j) = \cos(\phi_i + \phi_j) = \tilde{x}_i \tilde{x}_j - \sqrt{1 - (\tilde{x}_i)^2} \sqrt{1 - (\tilde{x}_j)^2}, \quad (4)$$

$$GADF(i, j) = \sin(\phi_i - \phi_j) = \tilde{x}_j \sqrt{1 - (\tilde{x}_i)^2} - \tilde{x}_i \sqrt{1 - (\tilde{x}_j)^2}. \quad (5)$$

The GAF matrix was constructed in the original representation space of the time series. Therefore, the encoding method has significant advantages in terms of efficiency, while avoiding

temporal information loss in the process of representation space transformation [25]. Work [25] pointed out that the GAF has two significant advantages. First, the GAF matrix preserves the time dependence of the time series, that is, each element of the GAF matrix was generated sequentially from the top left to the bottom right according to the temporal order of the original time series. Second, the GAF matrix contains temporal correlations; for example, $GAF(i, j)$ denotes the time interval $k = |i - j|$ relative correlation in the direction.

In practical applications, direct encoding of the bearing vibration signal in the time domain into a GAF matrix is often unsatisfactory. This is because the bearing signal in the time domain can easily be contaminated by noise. Therefore, we encoded the amplitude-spectrum sequence of the bearing signal into the GAF image. In the amplitude spectrum, the energy of useful information is concentrated in a narrow range of frequency bands, whereas the noise energy is distributed over the entire frequency band. Assuming that $\bar{X} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ is the amplitude spectrum of a bearing vibration signal, such that \bar{X} can be denoised according to Eq. (6):

$$(\bar{x}_i)' = \begin{cases} \bar{x}_i - \mu, & \bar{x}_i > \mu, \\ 0, & \bar{x}_i \leq \mu, \end{cases} \quad (6)$$

where, μ is the mean value of X . As the amplitude spectrum of the vibration signal is symmetric, it is possible to consider only the left amplitude spectrum. Fig. 2 and 3 show the GASF and GADF images of 10 different bearing faults, respectively. The class distributions of these faults are shown in Table 1.

Table 1. Class distribution of 9 types of faults

Fault type	Load (HP)	Speed (rpm)	Ball	Inner race	Outer race
0.007"	0	1797	class 1	class 4	class 7
0.014"	0	1797	class 2	class 5	class 8
0.021"	0	1797	class 3	class 6	class 9

3. Denoising method

In real applications, the sampling points of the time series are usually very large; therefore, it is necessary to reduce the dimensionality of the time series before GAF encoding. Considering that the piecewise aggregate approximation (PAA) algorithm [28] can not only preserve the basic trend of the time series but also has low time complexity, we use the PAA algorithm to pre-process the bearing vibration signal.

PAA is a simple and effective time series smoothing algorithm that preserves the trend of the time series. The time complexity of PAA is low; therefore, the PAA algorithm is widely used in time series analysis problems. Considering the time series $X = \{x_1, x_2, \dots, x_n\}$ is mapped to a new time series $\hat{X} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m\}$, \hat{x}_i can be calculated using the following formula:

$$\hat{x}_i = \frac{m}{n} \times \left(\sum_{j=n/m(i-1)+1}^{n/(m \times i)} x_j \right), \quad 1 \leq i \leq m. \quad (7)$$

From Eq. (7), we find that X is sequentially divided into m blocks of equal size, and the mean value of each block is used to re-represent the block. The PAA algorithm has a certain noise reduction effect, as it uses the mean value to smooth the data. Clearly, the selection of m is crucial. If m is too large, the smoothed result loses the original structural information, where m is too small, and the effect is not suitable for noise reduction. From Eq. (7), we can also see that the traditional PAA needs to satisfy n/m as an integer. For n/m is a non-integer number that can be found in [29] and [30].

3.1. Stacking integration methodology

Bagging, boosting, and stacking methods are three commonly used ensemble learning methods. The bagging method is an algorithm to reduce the variance in the estimate by using voting or mean reversion to achieve the fusion of multiple decision results [31]. Boosting method can upgrade weak learners to strong learners. Unlike the parallel learning approach of the bagging method, boosting method is a sequential framework. Boosting method works by sequentially training an initial learner from the training set, and then adjusting the distribution of training samples according to the results of the initial learner, thus, those instances of wrong decisions of the previous initial learner will be received attention. AdaBoost [32] method is a very classic boosting method. The stacking method is a different fusion method that is essentially representation learning. The principle of the stacking method is that they perform the second-stage learning with the result of the initial learner. Stacking method has yielded unusually brilliant results in many data mining competitions (e.g., data science competitions on the Kaggle platform). For example, in the solution proposed by the grand prize winner of the 2009 Netflix recommendation competition, integrating multiple initial learners is the core of its design [34].

In the case of the classification task, the basic process of the stacking method is to learn different classification algorithms $\mathcal{L}_1, \dots, \mathcal{L}_K$ on the dataset D . $d_i = (x_i, y_i) \in D$ is an instance, where x_i is the feature vector and y_i is the corresponding label. In the first stage of the stacking method, a set of base classifiers C_1, \dots, C_K where $C_i = \mathcal{L}_i(D)$ are generated. In the second stage, a meta-classifier was learned based on the outputs of the base classifiers. Note that the leave-one-out method or cross-validation [35] was applied to generate the training set for learning the meta-classifiers [33]. For the leave-one-out method, each base classifier uses almost all examples and leaves the remaining one for testing. The procedure can be formalized as $\forall i = 1, \dots, n$ (n is the number of examples), $C_t^i = \mathcal{L}_t(D - d_i)$, $\forall t = 1, \dots, K$, and next, the base learner is used to classify d_i by $\tilde{y}_i^t = C_t(x_i)$. Therefore, d_i can be reconstructed to a new vector $((\tilde{y}_i^1, \dots, \tilde{y}_i^K), y_i)$. The inputs of the meta-learning phase comprised the predictions of the base classifier.

As the leave-one-out method reconstructs only a sample per learning, it increases the time cost of the reconstruction step, whereas cross-validation predicts a pre-defined subset of the original sample set at a time and gets the predictions of the base classifier on these subsets. Thus, cross-validation is preferred for applying the stacking algorithm for big data.

4. Experiments

Here, we used CWRU bearing data [36] to verify the effectiveness of the proposed method. The dataset comprised a multivariate vibration time series generated by the bearing test equipment, as shown in Fig. 4. In this study, the bearing dataset included the following four conditions: normal, outer race failure, inner race failure, and roller failure. The drive-side vibration signal was used with a sampling rate of 48 kHz and a motor load of 0 hp. Table 1 shows the class distributions of the selected data.

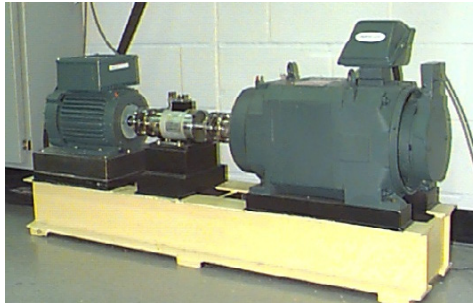


Fig. 4. Test-stand [36]

Table 2. Accuracy rate of 5 methods on raw CWRU

Algorithm	GASF+2D-CNN	GADF+2D-CNN	WPI+2D-CNN	Amplitude spectrum + 1D-CNN	Ensemble method
acc	99.4 %	99.8 %	98.7 %	75.6 %	100 %

5. Experimental design

Here, we used a 2D-CNN as the initial learner. Considering useable data is small, the 2D-CNN model includes only three convolutional layers, three pooling layers, and one fully connected structure. Meanwhile, the batch normalization [37] and the dropout [38] methods are used to reduce the risk of deep model over-fitting. The 2D-CNN uses the cross-entropy loss function, and Adam optimizes the algorithm [39]. We used the random forest regression method as the meta-learner. We follow the common practice of dividing the CWRU dataset was divided into the training set, test set, and validation set in the ratio of 0.5, 0.3, and 0.2. The training set was used to train the 2D-CNN and the validation set was used to train the random regression model. The experiments are repeated 10 times under different random seeds, and the final results are taken as the mean value of the 10 experiments. Considering the uniform distribution of classes, performing different algorithms can be well measured by the traditional classification accurate rate:

$$acc = \frac{\text{Correctly Identified Instances}}{\text{Total Instances}} \times 100\%, \quad (8)$$

5.1. Raw CWRU data

In this section, the experiments were divided into two parts. The first was to verify the validity of the image interpretation of vibration signals. Its purpose was to compare the fault recognition performance before and after image encoding. We used the 1D-CNN model to learn one-dimensional vibration signals and applied the 2D-CNN model to the imaged data. Second, we compare the performance of the proposed method with that of the existing method in [2] (hereafter referred to as WPI+2D-CNN). The topological structure of 2D-CNN is described in the above section. In addition, the 1D-CNN model contains a one-dimensional convolutional layer, a pooling layer, and a fully connected structure. To reduce the overfitting of CNN, batch normalization and dropout are used to reduce the risk of overfitting.

We showed the accuracy rates on the original CWRU data in Table2. From the table, we can conclude that: 1) the vibration signal encoded to a two-dimensional image can be better learned to achieve better performance; 2) the proposed ensemble method gets better performance than the existing method. Although WPI+2D-CNN also gets the accurate rate close to our approach, it has a high time cost than the proposed method. Due to GAFs performing the outer product of the time series, so the time complexity of GAFs is $O(N^2)$, N is the length of time series. Work [2] does not give the time complexity of WPI, such that we add a test to compare their runtimes. Table 3 shows the time consumption of imaging the raw CWRU data using GAFs and WPI respectively. From the table, we can see that WPI costs 2938.267 seconds, which is awful larger than GASF and GADF.

Table 3. Runtime of GAFs and WPI

	GASF	GADF	WPI
Runtime (s)	2.289 (s)	2.236 (s)	2938.267 (s)

Why can fusing the knowledge of both GASF and GADF can improve accurate rates? From Fig. 2, we can see that class 5, class 6, and class 7 have a similar representation in the GASF image while presenting a significant difference in the GADF domain. In the same light, class 2 and class 3 have similar GADF features, instead, have different GASF features. Fig. 2 and 3 explain the advantages of our method. In fact, from the perspective of information theory, fusing different

representation features of the GAF can increase the information entropy of the inputs and help improve the accuracy of the learning-based prediction model.

5.2. Noise-added CWRU data

To further verify the robustness of our method, we added Gaussian white noise to raw CWRU data. We added noise to the vibration signal with SNR of $[-5 \text{ dB}, -2 \text{ dB}, 2 \text{ dB}, 5 \text{ dB}]$. The experimental results are in Table 4. From the table, we can see that: 1) the proposed method can achieve better performance than WPI+2D-CNN in a noisy environment. 2) The fault identification performance of GADF+2D-CNN is close to GASF+2D-CNN. 3) The ensemble model fusing the GAF image information can achieve excellent performance. Note that for the bearing signal with complex noise, we can choose advanced denoising methods instead of formula (6). Examples include deep learning, wavelet shrinkage-based, SVD-based methods, and EMD-based methods.

In summary, we can conclude that fusing the knowledge of the GASF and the GADF can improve the performance of bearing fault diagnosis for the CWRU dataset. Considering that the time cost of building the GAF was low, the proposed method was more in line with the practical application requirements.

Table 4. Accuracy rate of 5 methods on noisy CWRU

	-5 dB	-2 dB	2 dB	5 dB
WPI+2D-CNN	87.1 %	93.2 %	97.6 %	97.8 %
GASF+2D-CNN	99.0 %	99.0 %	99.1 %	99.4 %
GADF+2D-CNN	99.4 %	99.1 %	99.0 %	99.6 %
Ensemble method	99.8 %	100 %	100 %	100 %

6. Conclusions

This study proposed a bearing fault detection method that combines im-age-interpreting vibration signals and integrating deep learning, which can realize the accurate identification of bearing faults. Our method encoded one-dimensional vibration signals into a two-dimensional image and then used a 2D-CNN to obtain the initial decision result. Finally, we introduced a decision layer integration method to realize the fusion of multiple underlying decisions. Experiments on the CWRU real-world dataset show that the proposed method can obtain a better recognition accuracy rate than the existing method (i.e., WPI+2D-CNN), even when Gaussian white noise is added to the original vibration signal.

Altogether, the learning-based method for bearing fault detection is provided in this work. Next, we plan to apply our method to different publicly available bearing failure datasets and laboratory datasets.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 61901191), and the Shandong Provincial Natural Science Foundation (Grant No. ZR2020LZH005).

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mechanical Systems and Signal Processing*, Vol. 100, No. 2, pp. 439–453, Feb. 2018, <https://doi.org/10.1016/j.ymssp.2017.06.022>
- [2] 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, <https://doi.org/10.1109/tim.2017.2674738>
- [3] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics – a review," in *2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*, Aug. 2019, <https://doi.org/10.1109/demped.2019.8864915>
- [4] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Transactions on Instrumentation and Measurement*, Vol. 66, No. 7, pp. 1693–1702, Jul. 2017, <https://doi.org/10.1109/tim.2017.2669947>
- [5] W. Bao, X. Tu, Y. Hu, and F. Li, "Envelope spectrum L-kurtosis and its application for fault detection of rolling element bearings," *IEEE Transactions on Instrumentation and Measurement*, Vol. 69, No. 5, pp. 1993–2002, May 2020, <https://doi.org/10.1109/tim.2019.2917982>
- [6] M. Chen, D. Yu, and Y. Gao, "Fault diagnosis of rolling bearings based on graph spectrum amplitude entropy of visibility graph," (in Chinese), *Journal of Vibration and Shock*, Vol. 40, No. 4, pp. 23–29, 2021, <https://doi.org/10.13465/j.cnki.jvs.2021.04.004>
- [7] Z. Zhao and S. Yang, "Sample entropy-based roller bearing fault diagnosis method," (in Chinese), *Journal of Vibration and Shock*, Vol. 31, No. 64, pp. 23–29, 2021, <https://doi.org/10.13465/j.cnki.jvs.2012.06.012>
- [8] C. Liu et al., "Rolling bearing fault diagnosis based on variational mode decomposition and fuzzy C-means clustering," *Proceedings of the Chinese Society of Electrical Engineering*, Vol. 35, No. 13, pp. 1–8, Aug. 2016.
- [9] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, Vol. 65, No. 7, pp. 5990–5998, Jul. 2018, <https://doi.org/10.1109/tie.2017.2774777>
- [10] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, Vol. 93, pp. 490–502, Nov. 2016, <https://doi.org/10.1016/j.measurement.2016.07.054>
- [11] I. H. Ozcan, O. C. Devecioglu, T. Ince, L. Eren, and M. Askar, "Enhanced bearing fault detection using multichannel, multilevel 1D CNN classifier," *Electrical Engineering*, Vol. 104, No. 2, pp. 435–447, Apr. 2022, <https://doi.org/10.1007/s00202-021-01309-2>
- [12] J. Cao, S. Wang, X. Yue, and N. Lei, "Rolling bearing fault diagnosis of launch vehicle based on adaptive deep CNN," (in Chinese), *Journal of Vibration and Shock*, Vol. 39, No. 5, pp. 97–104, 2020, <https://doi.org/10.13465/j.cnki.jvs.2020.05.013>
- [13] S. Dong, X. Pei, W. Wu, B. Tang, and X. Zhao, "Rolling bearing fault diagnosis method based on multilayer noise reduction technology and improved convolutional neural network," *Journal of Mechanical Engineering*, Vol. 57, No. 1, p. 148, 2021, <https://doi.org/10.3901/jme.2021.01.148>
- [14] G. Jin, "Research on end-to-end bearing fault diagnosis based on deep learning under complex conditions," University of Science and Technology of China, Hefei, 2020.
- [15] S. Haidong, J. Hongkai, L. Xingqiu, and W. ShuaiPeng, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowledge-Based Systems*, Vol. 140, No. 1, pp. 1–14, Jan. 2018, <https://doi.org/10.1016/j.knsys.2017.10.024>
- [16] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Transactions on Instrumentation and Measurement*, Vol. 67, No. 1, pp. 185–195, Jan. 2018, <https://doi.org/10.1109/tim.2017.2759418>
- [17] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, Vol. 240, No. 3, pp. 98–109, May 2017, <https://doi.org/10.1016/j.neucom.2017.02.045>
- [18] H. Jiang, X. Li, H. Shao, and K. Zhao, "Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network," *Measurement Science and Technology*, Vol. 29, No. 6, p. 065107, Jun. 2018, <https://doi.org/10.1088/1361-6501/aab945>

- [19] M. Landauskas, M. Cao, and M. Ragulskis, "Permutation entropy-based 2D feature extraction for bearing fault diagnosis," *Nonlinear Dynamics*, Vol. 102, No. 3, pp. 1717–1731, Nov. 2020, <https://doi.org/10.1007/s11071-020-06014-6>
- [20] H. Wang, J. Xu, R. Yan, and R. X. Gao, "A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN," *IEEE Transactions on Instrumentation and Measurement*, Vol. 69, No. 5, pp. 2377–2389, May 2020, <https://doi.org/10.1109/tim.2019.2956332>
- [21] X. Zhu, J. Zhao, D. Hou, and Z. Han, "An SDP characteristic information fusion-based CNN vibration fault diagnosis method," *Shock and Vibration*, Vol. 2019, p. 3926963, Mar. 2019, <https://doi.org/10.1155/2019/3926963>
- [22] H. Wang, J. Xu, R. Yan, C. Sun, and X. Chen, "Intelligent bearing fault diagnosis using multi-head attention-based CNN," *Procedia Manufacturing*, Vol. 49, pp. 112–118, 2020, <https://doi.org/10.1016/j.promfg.2020.07.005>
- [23] Y. Xu, Z. Li, S. Wang, W. Li, T. Sarkodie-Gyan, and S. Feng, "A hybrid deep-learning model for fault diagnosis of rolling bearings," *Measurement*, Vol. 169, p. 108502, Feb. 2021, <https://doi.org/10.1016/j.measurement.2020.108502>
- [24] D. Neupane, Y. Kim, and J. Seok, "Bearing fault detection using scalogram and switchable normalization-based CNN (SN-CNN)," *IEEE Access*, Vol. 9, pp. 88151–88166, 2021, <https://doi.org/10.1109/access.2021.3089698>
- [25] Z. Wang and O. Tim, "Imaging time-series to improve classification and imputation," in *Proceedings of the 24th International Conference on Artificial Intelligence Marina del Rey*, pp. 3939–3945, 2015.
- [26] L. Breiman, "Stacked regressions," *Machine Learning*, Vol. 24, No. 1, pp. 49–64, Jul. 1996, <https://doi.org/10.1007/bf00117832>
- [27] D. H. Wolpert, "Stacked generalization," *Neural Networks*, Vol. 5, No. 2, pp. 241–259, Jan. 1992, [https://doi.org/10.1016/s0893-6080\(05\)80023-1](https://doi.org/10.1016/s0893-6080(05)80023-1)
- [28] Z. Zhu et al., "Time series mining based on multilayer piecewise aggregate approximation," in *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, pp. 174–179, Jul. 2016, <https://doi.org/10.1109/icalip.2016.7846629>
- [29] J. Lin et al., "Experiencing SAX: a novel symbolic representation of time series," *Data Mining and Knowledge Discovery*, Vol. 15, No. 2, pp. 107–144, Jul. 2016.
- [30] Y. Huang, W. Jin, P. Ge, and B. Li, "Radar emitter signal identification based on multi-scale information entropy," (in Chinese), *Journal of Electronics and Information Technology*, Vol. 41, No. 5, pp. 1084–1091, 2019, <https://doi.org/10.11999/jeit180535>
- [31] L. Breiman, "Bagging predictors," *Machine Learning*, Vol. 24, No. 2, pp. 123–140, Aug. 1996, <https://doi.org/10.1007/bf00058655>
- [32] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, Vol. 55, No. 1, pp. 119–139, Aug. 1997, <https://doi.org/10.1006/jcss.1997.1504>
- [33] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?," *Machine Learning*, Vol. 54, No. 3, pp. 255–273, Mar. 2004, <https://doi.org/10.1023/b:mach.0000015881.36452.6e>
- [34] J. Sill, T. Gabor, M. Lester, and L. David, "Feature-weighted linear stacking," *ArXiv Preprint, ArXiv:0911.0460*, 2009.
- [35] Z. Zhou, *Machine Learning*. Beijing, China: Tsinghua Press, 2016.
- [36] "Bearing data center." Case Western Reserve University. <https://csegroups.case.edu/bearingdatacenter/pages/download>
- [37] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, pp. 448–456, 2015, <https://doi.org/10.48550/arxiv.1502.03167>
- [38] Srivastava N. et al., "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958, 2014, <https://doi.org/10.5555/2627435.2670313>
- [39] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *ArXiv Preprint, ArXiv:1412.6980*, 2014, <https://doi.org/10.48550/arxiv.1412.6980>



Yanfang Han received the B.S. degree in highway and railway engineering from Southwest Jiaotong University, Chengdu, China, in 2009, and an M.S. degree in geologic engineering from Southwest Jiaotong University, Chengdu, China, in 2013. She is currently a lecturer at the Sichuan College of Architectural Technology. Her research interests include machine learning, fault detection, and landslide monitoring.



Baozhu Li received the Ph.D. degree in computer science and technology from Beijing University of Posts and Telecommunications, Beijing, China, in 2017. From 2013 to 2014, he was a Visiting Student with the Broadband Communications Research (BBCR) Lab, Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is currently with the Internet of Things and Smart City Innovation Platform, Zhuhai Fudan Innovation Institute. His research interests include vehicular networks, mobile edge computing, machine learning, and resource management.



Yingkun Huang received the B.S. degree in computer science and technology from Southwest University, Chongqing, China, in 2011, and the M.S. degree in electric engineering from Huaqiao University, Xiamen, China, in 2015. He is currently pursuing the Ph.D. degree in Electrical Engineering at Southwest Jiaotong University, Chengdu, China. His research interests include machine learning, data mining, signal processing, and knowledge discovery.



Liang Li received the B.S. degree from the Nanjing University of Information Science and Technology, Jiangsu, China, in 2013, and the M.S. degree in electric engineering from Sichuan University of Science and Engineering, Chengdu, China, in 2017. He is currently pursuing the Ph.D. degree in Electrical Engineering at Southwest Jiaotong University, Chengdu, China. His research interests include deep learning and machine learning.