Iranian Society of Acoustics and Vibration
The 13th ISAV2023 International Conference on Acoustics and Vibration
20 , 21 Dec 2023 Tehran - Iran

Fault Diagnosis of Rotating Machines based on the Enhanced Multi-Scale Convolutional Neural Network Approach

Raha Pedram [a], Ali Chaibakhsh [a*]

[a] *Faculty of Mechanical Engineering, Intelligent Systems and Advanced Control Lab, University of Guilan, 41996-13776, Rasht, Iran.*
[*] *chaibakhsh@guilan.ac.ir*

Abstract

The fault diagnostics of rotating machinery significantly affect the dependability and safety of modern industrial systems. Advanced fault diagnosis techniques have taken over the challenging and uncertain process of human analysis, boosting the effectiveness of fault diagnosis. The accuracy of fault diagnosis is enhanced by employing deep learning models, well-known for their outstanding ability to process complex relationships across multiple layers. This paper introduces a novel approach to analysing rotating faulty machines using a multi-scale convolutional neural network (MSCNN) model with an attention mechanism layer. Different faults including unbalanced and bearing faults are investigated based on the pre-processed signals, which are processed using the window-sliding and short-time Fourier transform methods with 13 distinct modes. The proposed approach offers a cost-effective solution without compromising reliability by leveraging a subset of three sensors out of six. The multi-scale CNN model, featuring different kernel sizes, simultaneously captures local and global features. Experimental evaluations demonstrate the effectiveness of the proposed approach in fault classification with an accuracy of 99.68%, which shows the feasibility and accuracy of the proposed technique.

Keywords: Fault diagnosis; Convolutional Neural Networks (CNNs); Unbalanced; Bearing fault.

1. Introduction

Rotating machines play a vital role in prime mover equipment and act as essential assets in the industrial realm. The failure of a single component can lead to severe losses that have profound implications for production, quality control, financial stability, and personnel safety. To mitigate such risks and ensure continuous operations within the industry, it becomes imperative to employ preventive measures, such as regular maintenance and timely restorations. By implementing advanced fault diagnosis methods, the ability to identify various faults in rotating machinery can bring about significant transformations in the industrial sector [1]. Vibration analysis is one common technique for identifying issues with rotating machinery, such as mass imbalance and bearings faults. The fault diagnosis process entails collecting data, identifying faults, and extracting essential features. By implementing techniques that preserve the integrity of vibration characteristics, we can substantially enhance the efficacy of the fault diagnosis process [2]. Conventional signal processing methods, such as support vector machines (SVMs) and fuzzy inference, have long been employed in fault detection in rotating machinery and industrial settings. However, as industrial challenges grow increasingly complex, these methods have demonstrated limitations, particularly in handling signals associated with nonlinear problems. Their suboptimal outcomes have prompted the exploration of alternative approaches [3]. Deep learning models, with their powerful nonlinear functions, offer a promising solution for fault detection tasks. They excel at understanding complex data patterns, outperforming traditional methods [4].

Researchers have explored diverse approaches for diagnosing faults in rotating machinery in recent years. Zhang et al. [5] addressed high-frequency noise in rotary machinery fault diagnosis by incorporating a wide kernel in the first layer of their CNN. Their efforts led to the development of the convolution model with Wide First-Layer Kernels designed for accurate fault diagnosis in rotary machinery. Using 1D raw time series data as inputs in the CNN model for motor fault detection, as described in [6], eliminates the requirement for laborious feature extraction processes, resulting in more efficient and effective detection. Wei Zhang et al. [7] introduced a Convolutional Neural Network (CNN) architecture comprising two convolutional layers to efficiently identify and diagnose faults in bearings using substantial training data. The authors of [8] presented a deep learning method for bearing defect diagnosis using the short-time Fourier transform (STFT) and the LAMSTAR neural network. Experimental validation verifies its efficacy for various bearing faults and working conditions. In [9], a novel method for early diagnosis of gear pitting problems in different working environments was presented. The method employed a novel adaptive 1D separable convolution network with residual connections to classify defects accurately with a significantly reduced set of model parameters compared to standard convolutional neural networks. In [10], a sparse deep learning technique called SDSN was introduced to address the overfitting risk in deep networks. SDSN used sparsity in output label encoding and a sparse regularization term to enhance fault classification performance.

Despite several studies demonstrating satisfactory accuracy in fault detection, the demand for more effective models in this area is still compelling. This is primarily due to the significant financial consequences of repairs and maintenance. Therefore, developing a robust fault diagnosis model is crucial to accurately detect various faults in rotating machines, even in scenarios with limited sensors or data availability.

This paper presents a novel fault diagnosis framework for rotating machinery. We used a multi-scale convolutional neural network (MSCNN) with a custom attention layer to improve diagnostic accuracy. Our approach captured fine and coarse details for comprehensive fault pattern analysis. Unlike previous research, we considered a wider range of 13 distinct conditions, including six bearing faults and seven unbalanced faults (single and combined). We employed the Short-Time Fourier Transform (STFT) for pre-processing and achieved impressive robustness and accuracy using only three out of six available sensors, keeping computational costs low. This approach is practical for industrial fault diagnosis in rotating machinery, reducing financial burdens.

Section 2 delves into the preliminaries of this research. Section 3 provides an overview of the proposed method. Subsequently, Section 4 presents a comprehensive analysis of the experimental results. Finally, in Section 5 , the results are concluded, and the potential future directions for further research are outlined.

2. Preliminaries

2.1 Short-Time Fourier Transform (STFT)

The fundamental concept of the Short-Time Fourier Transform (STFT) is to identify the frequency components of a signal by applying a time-limited window function to the signal before performing the Fourier transform. By systematically moving and overlapping the window along the time axis, the signal can be analysed segment by segment, resulting in frequency domain representations at various time intervals. This enables the examination of evolving frequency characteristics of the signal over time [3].

$$S_i(w) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jwt} s(\tau) h(\tau - t) d\tau \tag{1}$$

In Eq. (1), the window function $h(\tau)$ plays a key role in shaping time and frequency resolutions. The under-analysis signal is denoted as $s(\tau)$. By adjusting the window length, the Short-Time Fourier Transform (STFT) allows for a flexible trade-off between accuracy in time and frequency [11]. The frequency scale interval is designed to align perfectly with the frequency resolution. The resolution, in turn, is derived by dividing the sampling frequency by the number of samples, with a specific sampling frequency of 10 kHz used in this case.

2.2 A Concise Overview of Convolutional Neural Network

2.2.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) utilize a signal convolution process with filters (kernels) to extract features from input signals. This technique allows for pattern recognition regardless of the spatial position of the data. The kernels are fine-tuned during supervised training to optimize parameters and evaluate results. The final characteristic map consists of multiple layers with weights and biases but similar structures. By activating these units, the convolutional operation is applied to the entire input dataset, producing the expected convolutional output [12].

$$y_i^{(l+1)}(j) = K_i^l * x^l(j) + b_i^l \tag{2}$$

The notations K_i^l and b_i^l represent the weights and bias related to the *i*-th filter kernel in layer l. Similarly, the *j*-th local region within layer l is called $x^l(j)$. In addition, $y_i^{(l+1)}(j)$ represents the input received by the *j*-th neuron in frame *i* of layer *l* + 1 during the convolution process [5].

2.2.2 Parametric rectified linear unit

The activation layer in a CNN is a vital component that determines neuron activation in the next layer through nonlinear operations. Widespread activation functions such as ReLU, sigmoid, and tanh are utilized, having widely adopted ReLU for its versatility in various CNN architectures. Additionally, ReLU is unique because it only activates positive neurons, resulting in fewer activated neurons compared to other activation functions. This particular attribute contributes to faster network convergence during training. The mathematical representation for the ReLU operation can be described by the following equation [13].

$$f(x) = \max(0, x) \tag{3}$$

2.2.3 Batch normalization

In order to speed up the training of neural networks, batch normalization (BN) standardizes the data in each batch. The internal covariance shift is decreased by applying it typically after the convolution layer and before the activation function. The ability of the network to generalize is increased by the input data normalization of the BN, which removes the need for the network to adapt to various data distributions [14].

2.2.4 Pooling Layer

CNNs typically have pooling layers added after the convolutional layer. These layers perform an important down-sampling process by decreasing the spatial dimensions of the collected features and optimizing the overall network parameters. The superior performance favours the max-pooling layer over other pooling methods. This layer reduces the network complexity by performing a local maximum on the input features. Additionally, it helps create location-independent features, which are crucial for building a CNN model that learns effectively and robustly. Adding max-pooling layers can simplify the network and improve its capacity to detect and express the key patterns of the data [5].

2.2.5 Drop out

The dropout operation in a neural network is critical for performance improvement. The operation is accomplished by random deactivation of specific elements, setting them to zero, and scaling the non-zero elements simultaneously. When an element is zeroed out, it is ignored. As a result, to create accurate predictions of the target variable, the model is forced to rely on the remaining information. The model learns using a subset of characteristics throughout each training cycle by introducing unpredictability through dropout. Due to this unpredictability, the model learns different combinations of characteristics, encouraging each feature to contribute significantly to the model predictions. This process prevents the model from relying too much on specific features, reducing the risk of overfitting [15].

3. The proposed method

In this paper, a multi-scale CNN model with an attention mechanism for rotating fault machinery to provide an in-depth exploration of vibration signal recordings associated with a diverse set of 13 fault types is introduced. Among these are mass imbalance faults in 6 rotating discs, combined imbalance faults involving discs 1 and 4, outer race faults in 3 bearings, and three ball bearings faults. Collectively, these fault types form a comprehensive dataset consisting of 13 distinct classes. The rotating machine, depicted in Figure 1. has six sensors; however, we specifically utilized only three. This selection specifically targets industries that face the challenge of limited sensor availability due to their increasingly high costs. This approach addresses the practical concerns and challenges faced in real-world applications, providing a cost-effective solution without compromising the reliability and accuracy of the model. Here, a window-sliding technique has been implemented as the initial pre-processing step on raw vibration signals, with a window size of 1500 and a stride of 1000. Subsequently, the signals were treated to the Short-Time Fourier Transform (STFT) algorithm during the initial pre-processing stage. The signals were examined in both the temporal and frequency domains, revealing detailed spectrum properties and identifying frequency components. The main goal of this research was to feed shuffled and normalized pre-processed data into a multi-scale neural network that was enhanced with an attention mechanism. This architecture simultaneously advances two feature extractor models. The initial feature extractor consists of two two-dimensional CNNs, employing a 3×3 kernel size.

In contrast, the second model follows a similar architecture but adopts a larger kernel size of 7 ×7. This strategic approach enables the simultaneous extraction of local and global features, capturing a comprehensive range of information within the input data. In this model, the max pooling was incorporated to extract key features and leverage the ReLU activation function. For the further enhancement of the performance, batch normalization was applied after each convolutional layer, which favours efficiency optimization.

The attention mechanism is a resource allocation scheme that helps manage information overload in situations with limited computational power. It ensures that important data is processed by assigning attention weights to input features [16]. In the present attention layers, initially, these weights are set to 1, treating all elements equally. However, the model learns to modify the attention weights through training and backpropagation, allowing it to prioritize and assign distinct weights to different elements based on patterns and connections within the training data. This optimization of resource utilization enables the model to handle large amounts of information effectively. Attention scores are computed by multiplying the input data with the attention weights and summing the results. To ensure that the attention scores are normalized adequately within the range of [0, 1], a softmax activation function is applied to obtain attention weights, representing the relative importance of each element and indicating which elements should be emphasized or prioritized during computation. These attention weights are then applied element-wise to the input data, resulting in a weighted representation known as attention. The integration of attention within the final layer allows the model to selectively focus on relevant information, enhancing its ability to process sequential data effectively.

The second model employs a similar design with a larger kernel, applying 7×7 kernel size. These unique feature maps from the two models are concatenated and flattened to create a 1D vector. Three dense layers are added sequentially, with a dropout of 0.2 applied after the second dense layer to randomly set a fraction of input units to 0 during training, helping to prevent overfitting. The first two dese layers utilize the rectified linear unit (ReLU) activation function, and the last dense output layer produces a probability distribution over 13 classes, with 13 units and softmax activation. The tests were carried out using the Tensorflow toolbox [17]. Finally, we used the Adam stochastic optimization algorithm, renowned for its effectiveness in handling large input data sets and models with numerous parameters, to minimize the loss function. The input shape would take the form of (None, 12, 129, 3). Here, the second dimension represents the temporal intervals, the third dimension signifies the frequency bins, and the fourth dimension corresponds to the number of sensors utilized. The model was trained on 60% of the data, comprising 9397 samples. 20% of the remaining was used for testing, and another 20% was reserved for validation. Figure 2. visually represents the sequence of operations in the proposed scheme where N is time intervals, W represents frequency bins, and C is the number of sensors. Table 1. provides the specific parameters for the proposed model.

4. Results and Discussion

4.1 Experimental Setups

Utilizing a rotating machine simulation device proves highly effective in troubleshooting various industrial rotating machines. The device allows for replicating real-world faults under controlled conditions, enabling the application of different fault-finding techniques to address these issues. By simulating and resolving such faults, this device enhances troubleshooting capabilities for a wide range of industrial applications. A device obtained from the Guilan University laboratory was implemented to simulate the defect of a rotating machine and provide the vibration data used in this study. The device has a 370-watt electro-motor with a nominal speed of 2,825 rpm (In 50 Hz), and a CV100 Kinco® inverter controls the speed. On this apparatus, various bearing, unbalanced, and misalignment fault types can be simulated in various locations, and vibration data can be collected. Also, three PH204 bearings with FAG2302 ball bearings and six aluminum discs with 11, 12, 13, 14, 15, 16, and 17 centimeters diameters are installed on the system. Six accelerometers of model AD100T were used to collect vibration data. A data logger was used to receive the data from the sensors called the Advantech® USB 4711A and saved the data on a computer. The vibration signal was expertly captured in this article, with a precise sampling frequency of 10,000 Hz and a duration of 60 seconds.

Figure 1. The design of the rotating machine simulator

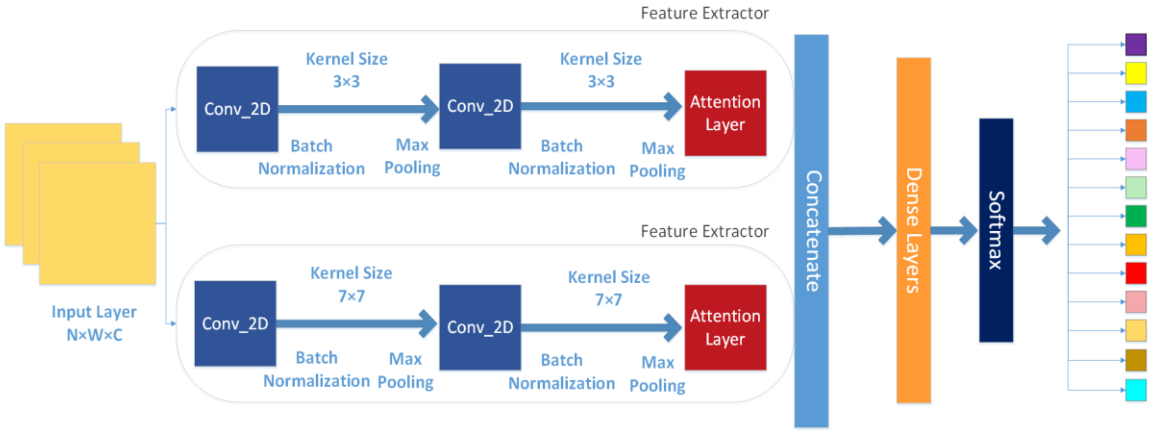

Figure 2. The architecture of the proposed method

4.2 Experimental Results

In order to validate the results presented of this work, they were compared to the other conventional and deep learning methods. Initially, the same dataset was used as input for a regular convolutional model, which had a similar architecture, the same as the first feature extractor with 3×3 kernel size with RMSprop optimizer and a learning rate of 0.01. However, it lacked the attention layer and the two accompanying feature extractors. Our proposed model outperformed the standard CNN in a thorough evaluation conducted over 50 epochs. Our model demonstrated remarkable performance from the initial epochs with minimal training loss that steadily approached zero by the fifth epoch. Additionally, our model achieved near-zero validation loss after 25 epochs and maintained this level throughout the rest of the training.

In contrast, the normal CNN model experienced significant fluctuations in validation loss beyond the 25th epoch, failing to achieve stable and satisfactory performance. These fluctuations ranged from approximately 0.2 to 0.8, making the traditional CNN model less reliable. Our model consistently and gradually decreased validation loss, ultimately converging to zero. Meanwhile, the proposed model has achieved an impressive validation accuracy of approximately 99%, demonstrating remarkable stability; however, the conventional model consistently fails to deliver reliable validation accuracy, exhibiting fluctuating results that vary unpredictably. Its performance can range from high to low with no guarantee of consistency. These observations establish the limitations of the traditional model in contrast to the robustness of the proposed approach. Utilizing MSCNN with an attention mechanism, the proposed approach exhibits superior effectiveness, demonstrated by its stable nature. In contrast, the limitations of the traditional model are evident through its fluctuating validation loss compared to the consistent performance of our approach.

Subsequently, a comparison was conducted between the proposed method and traditional approaches. Utilizing the Short Fast Fourier Transform, signal features were extracted using the Analysis of Variance (ANOVA) method. These features were input into common classifiers, which included a Support Vector Machine (SVM) with a C value set to 0.8, and a K-Nearest Neighbor (KNN) algorithm with K set to 5. However, as depicted in Table 2, the test accuracy percentages achieved by SVM and KNN were 90.45% and 84.42%, respectively, which fall below the acceptable threshold for model confidence.

The proposed approach was evaluated alongside Artificial Neural Network (ANN) methodology in the final stage. The results revealed that, overall, ANN exhibited inferior performance compared to convolutional methods, yielding a test accuracy of 89.84%. The multi-scale CNN with an Attention mechanism stood out among the compared methods, showcasing the highest level of performance with impressive test accuracy of **99.68%** and a validation accuracy of **99.78%.** Notably, when we compared our method to alternative approaches, all of them displayed validation accuracy rates in the range of 85% to 92%. This clearly underscores their performance and limited accuracy. The reported results are based on 25 repeated runs conducted on an ASUS K556U laptop equipped with an Intel ©Core i7 2.7GHz processor, utilizing Python 3.4.4 for all computations.

Table 1. Detailed parameters for the proposed CNN model architecture

Layers	Filters	Kernel Size	Output Size	Parameters
Input Layer	-	-	(None, 12, 129, 3)	0
Conv-2D	32	3×3	(None, 12, 129, 32)	896
Conv-2D	32	7×7	(None, 12, 129, 32)	4736
Batch Normalization	-	-	(None, 12, 129, 32)	128
Batch Normalization	-	-	(None, 12, 129, 32)	128
Max Pooling	-	2×2	(None, 6, 64, 32)	0
Max Pooling	-	2×2	(None, 6, 64, 32)	0
Conv-2D	64	3×3	(None, 6, 64, 64)	18496
Conv-2D	64	7×7	(None, 6, 64, 64)	100416
Batch Normalization	-	-	(None, 6, 64, 64)	256
Batch Normalization	-	-	(None, 6, 64, 64)	256
Max Pooling	-	2×2	(None, 3, 32, 64)	0
Max Pooling	-	2×2	(None, 3, 32, 64)	0
Attention Layer	-	-	(None, 3, 32, 64)	64
Attention Layer	-	-	(None, 3, 32, 64)	64
Concatenate	-	-	(None, 3, 32, 128)	0
Flatten	-	-	(None, 12288)	0
Dense Layer 1	-	-	(None, 64)	786496
Dense Layer 2	-	-	(None, 32), Dropout 0.2	2080
Dense Layer 3	-	-	(None, 13)	429
Total Parameters	-	-	-	914445

Table 2. Comparative accuracy results of the proposed method against other methods

Methods	Train Accuracy (60%)	Test Accuracy (20%)	Validation Accuracy (20%)
Traditional CNN	97.32%	91.63%	92.4%
Support Vector Machine	90.73%	90.45%	90.99%
K-nearest Neighbor	90.73%	84.42%	85.34%
ANN	98.2%	89.84%	89.85%
Proposed Method	**100%**	**99.68%**	**99.78%**

5. Conclusion

This research presents an innovative strategy for detecting faults in rotating machinery using a multi-scale convolutional neural network (MSCNN) model integrated with an attention mechanism. By analysing signal recordings and focusing on 13 distinct fault types, a cost-effective solution is offered, which is suitable for industries with limited sensor resources, employing only three of six available simulator sensors. By incorporating techniques such as window sliding, short-time Fourier transform (STFT) pre-processing, and a multi-scale neural network architecture featuring two feature extractors along with an attention mechanism in the last layer, the proposed model achieves an impressive test accuracy rate of **99.68**%. The performance surpasses that of the Normal CNN model, which has achieved an accuracy of 91.63%. Furthermore, our approach outperforms other examined techniques by a significant margin, providing valuable insights for developing more precise fault diagnosis methods in rotary machines. Our future work will focus on developing deep learning models specifically tailored to address the challenges posed by randomly occurring sensor failures. We aim to enhance the reliability and accuracy of sensor-based systems operating in dynamic and uncertain environments.

REFERENCES

1. 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, 2017.
2. M. Vishwakarma, R. Purohit, V. Harshlata, and P. Rajput, "Vibration analysis & condition monitoring for rotating machines: a review," *Materials Today: Proceedings,* vol. **4**, no. 2, pp. 2659-2664, 2017.
3. Y. Wang, M. Yang, Y. Li, Z. Xu, J. Wang, and X. Fang, "A multi-input and multi-task convolutional neural network for fault diagnosis based on bearing vibration signal," *IEEE Sensors Journal,* vol. **21**, no. 9, pp. 10946-10956, 2021.
4. H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mechanical Systems and Signal Processing,* vol. **102**, pp. 278-297, 2018.
5. W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors,* vol. **17**, no. 2, p. 425, 2017.
6. T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Transactions on Industrial Electronics,* vol. **63**, no. 11, pp. 7067-7075, 2016.
7. W. Zhang, G. Peng, and C. Li, "Rolling element bearings fault intelligent diagnosis based on convolutional neural networks using raw sensing signal," in *Advances in Intelligent Information Hiding and Multimedia Signal Processing: Proceeding of the Twelfth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Nov., 21-23, 2016, Kaohsiung, Taiwan, Volume* **2**, 2017: Springer, pp. 77-84.
8. M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Transactions on Industry Applications,* vol. **53**, no. 3, pp. 3057-3065, 2017.
9. X. Li, J. Li, C. Zhao, Y. Qu, and D. He, "Gear pitting fault diagnosis with mixed operating conditions based on adaptive 1D separable convolution with residual connection," *Mechanical Systems and Signal Processing,* vol. **142**, p. 106740, 2020.
10. C. Sun, M. Ma, Z. Zhao, and X. Chen, "Sparse deep stacking network for fault diagnosis of motor," *IEEE Transactions on Industrial Informatics,* vol. **14**, no. 7, pp. 3261-3270, 2018.
11. C. T. Alexakos, Y. L. Karnavas, M. Drakaki, and I. A. Tziafettas, "A combined short time fourier transform and image classification transformer model for rolling element bearings fault diagnosis in electric motors," *Machine Learning and Knowledge Extraction,* vol. **3**, no. 1, pp. 228-242, 2021.

12. F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. **16**, no. 1, p. 115, 2016.
13. H. Srivastava and K. Sarawadekar, "A depthwise separable convolution architecture for CNN accelerator," in *2020 IEEE Applied Signal Processing Conference (ASPCON)*, 2020: IEEE, pp. 1-5.
14. T. Cooijmans, N. Ballas, C. Laurent, Ç. Gülçehre, and A. Courville, "Recurrent batch normalization," *arXiv preprint arXiv:1603.09025*, 2016.
15. X. Li, X. Kong, Z. Liu, Z. Hu, and C. Shi, "A novel framework for early pitting fault diagnosis of rotating machinery based on dilated CNN combined with spatial dropout," *IEEE Access*, vol. **9**, pp. 29243-29252, 2021.
16. Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. **452**, pp. 48-62, 2021.
17. TensorFlow. "TensorFlow." https://www.tensorflow.org/ (accessed March 25, 2023.)