American International Journal of Scientific & Technical Research

ISSN 3144-2153 Official Website: https://aijstr.com

Adaptive Wavelet Denoising Combined with Improved Transformer for Industrial Equipment Vibration Fault Diagnosis

Javier Manuel Ortiz [a], Bram Jan van der Meer [b]

Corresponding Author. E-mail: javier.ortiz@upm.es

[a] *Department of Signal Theory and Communications, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, Madrid, 28040, Spain*
[b] *Department of Mechanical Engineering, Delft University of Technology, Mekelweg 2, Delft, 2628 CD, Netherlands*

A B S T R A C T

Vibration signal analysis serves as one of the most mainstream technical routes for condition monitoring and fault diagnosis of rotating machinery such as rolling bearings in modern industrial systems. In practical industrial scenarios, vibration signals collected by sensors are inevitably mixed with strong background noise generated by mechanical coupling, electromagnetic interference and complex working condition fluctuations. The weak fault impulse features characterizing early bearing damage are often submerged in noise, which brings great challenges to the accurate identification of fault categories. Traditional fault diagnosis methods rely heavily on manually designed feature extraction rules, and their generalization ability under variable working conditions is limited. In recent years, Transformer models based on self-attention mechanism have achieved remarkable performance in time series analysis tasks, but the vanilla Transformer still suffers from excessive computational overhead and interference from redundant background features when directly applied to vibration signal diagnosis. To solve the above limitations, this paper constructs an industrial vibration fault diagnosis framework integrating adaptive wavelet denoising and improved Transformer. The adaptive wavelet threshold denoising algorithm is firstly adopted to suppress high-frequency noise components in raw vibration signals and retain the transient impact characteristics related to bearing faults. Afterwards, an improved Transformer model with gated multi-head attention is proposed to realize adaptive feature screening and long-range dependency mining of denoised vibration sequences. The gated unit embedded before multi-head attention can automatically evaluate the importance of different signal segments, suppress the response of redundant noise features, and guide attention heads to focus on the time segments containing fault impulses. Comparative experiments are carried out on the public bearing fault dataset, and the diagnostic accuracy of the proposed method is compared with classic machine learning models, convolutional neural networks and original Transformer. The experimental results demonstrate that the proposed integrated framework achieves superior classification accuracy and stronger anti-noise robustness, especially for early weak bearing fault samples. This work provides a feasible technical reference for intelligent fault diagnosis of rotating machinery under noisy industrial environments.

Keywords: *wavelet denoising; adaptive threshold; Transformer; gated multi-head attention; vibration signal; fault diagnosis*

1. Introduction

With the continuous advancement of intelligent manufacturing and industrial internet technologies, online condition monitoring and automatic fault diagnosis of key rotating equipment have become essential guarantees for stable, safe and low-loss operation of modern production lines. Rolling bearings act as core supporting components of various rotating machinery, and local defects such as inner race wear, outer race peeling and rolling element damage will generate periodic impact components in vibration signals during equipment operation. Once the early minor faults cannot be detected in a timely manner, continuous deterioration of bearing damage may trigger sudden shutdown of the whole mechanical system and even cause major safety accidents

Corresponding author : Javier Manuel Ortiz
E-mail address: javier.ortiz@upm.es

and economic losses (Li et al., 2021). Therefore, the research on high-precision intelligent diagnosis technology for bearing weak faults under complex noise conditions has important engineering application value.

Traditional bearing fault diagnosis mainly includes spectrum analysis, envelope demodulation and other signal processing technologies. These methods require professional technicians to manually observe the frequency spectrum and extract fault-related characteristic frequencies, which puts forward high requirements for professional knowledge and practical experience of operators. When the signal-to-noise ratio is low and fault features are weak, the artificial feature extraction process is prone to failure (Wang et al., 2019). In the past decade, deep learning has greatly promoted the development of intelligent fault diagnosis. Convolutional neural networks are widely applied to extract local vibration features through receptive field sliding calculation, while recurrent neural networks such as LSTM and GRU are used to model time-series dependence of vibration signals. Nevertheless, CNN is limited by fixed local receptive field and weak ability to capture long-distance correlation of signal sequences. Recurrent neural networks adopt stepwise sequential calculation, and the attenuation of long-distance information will reduce the diagnosis performance for weak fault samples (Zhao et al., 2020).

Transformer architecture based on self-attention mechanism has broken the limitation of sequential calculation of recurrent neural networks. It can directly calculate the correlation between any two sampling points within the whole sequence, which brings new research ideas for vibration time-series analysis (Vaswani et al., 2017). However, the original Transformer has obvious defects when directly applied to vibration fault diagnosis. The model will allocate attention calculation resources to all sampling points indiscriminately, and the residual noise and redundant normal background signals will interfere the extraction of fault features, especially under low signal-to-noise ratio working conditions (Zhang et al., 2022). Many existing studies simply stack Transformer encoder layers, lacking targeted feature screening mechanism oriented to vibration signal characteristics. In order to reduce the interference of redundant features and improve the model's ability to capture weak impact features, this paper introduces learnable gated units before multi-head attention calculation and constructs gated multi-head attention. Combined with adaptive wavelet denoising preprocessing, an integrated fault diagnosis framework is formed.

The main contributions of this paper are summarized as follows. First, an adaptive wavelet threshold denoising strategy is adopted for raw vibration signal preprocessing, which can adaptively adjust denoising threshold according to signal noise distribution, effectively suppress background noise while retaining transient fault impact features. Second, an improved Transformer model with gated multi-head attention is proposed. The embedded gated unit realizes adaptive screening of vibration sequence features before attention calculation, reduces redundant feature interference and enhances the model's recognition capability of weak fault features submerged in noise. Third, comparative experiments are carried out on the widely used rolling bearing fault dataset. The experimental results verify that the proposed method outperforms multiple mainstream deep learning models in diagnosis accuracy and anti-noise robustness, which confirms the effectiveness of the improved model structure.

The rest of this paper is arranged as follows. Section 2 introduces the principle of adaptive wavelet denoising for vibration signals. Section 3 elaborates the structure and working mechanism of the improved Transformer model with gated multi-head attention in detail. Section 4 describes the experimental dataset, evaluation metrics and experimental settings. Section 5 analyzes the experimental results and carries out comparative discussion. Section 6 summarizes the whole work and puts forward the prospect of follow-up research.

2. Adaptive Wavelet Denoising Method

In the actual industrial measurement process, the vibration signal collected by acceleration sensors is mixed with multiple types of noise components. The noise sources include random mechanical vibration generated by other rotating parts of the equipment, electromagnetic interference introduced by sensor transmission lines, and white noise generated by the acquisition circuit itself. Such noise components often overlap with the frequency band of early fault impact features, which makes it difficult to separate fault characteristic components by simple low-pass or high-pass filtering. Wavelet transform can decompose the original signal into wavelet coefficients of different frequency scales, which provides a feasible way to separate noise and fault impact features (Donoho & Johnstone, 1994). Different from fixed threshold denoising, adaptive wavelet threshold denoising can dynamically adjust

© 2026 Javier Manuel Ortiz. This open-access manuscript is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits unrestricted reuse, distribution, and reproduction in any medium, provided the original author and source are fully credited. Full license text: https://creativecommons.org/licenses/by/4.0/

threshold parameters according to the statistical characteristics of wavelet coefficients of each layer, so as to avoid over-denoising which may filter out weak fault impact features, or insufficient denoising caused by fixed threshold.

The overall process of adaptive wavelet denoising is described as follows. First, the raw one-dimensional vibration signal is decomposed into multiple layers by wavelet decomposition. After decomposition, two types of wavelet coefficients can be obtained: low-frequency approximation coefficients and high-frequency detail coefficients. The low-frequency approximation coefficients mainly reflect the slow-changing trend component of the signal, while the high-frequency detail coefficients contain transient impact features and noise components. The noise is mostly distributed in the high-frequency detail coefficients. Next, the adaptive threshold of each layer of high-frequency wavelet coefficients is calculated according to the noise variance estimated from the signal. The threshold value of each decomposition layer is adjusted independently instead of adopting a unified threshold for all layers. The soft threshold function is adopted to process the high-frequency wavelet coefficients. The coefficient values smaller than the threshold are suppressed, and the coefficients larger than the threshold are reserved and appropriately shrunk, which can reduce the oscillation phenomenon generated by hard threshold processing. Finally, wavelet reconstruction is performed using the processed high-frequency coefficients and the original low-frequency approximation coefficients, so as to obtain the denoised vibration signal.

The selection of wavelet basis function is a key step affecting the denoising effect. For bearing fault vibration signals with transient impact characteristics, wavelet basis with good compact support and vanishing moment characteristics is selected in this paper. The wavelet basis function with similar waveform to fault impact pulse can better match the transient fault characteristics in vibration signals, so that the fault impact components can be retained in the high-frequency coefficients during decomposition, while random noise is suppressed in the threshold processing stage. In the denoising process, the noise variance estimation is updated in a layer-by-layer manner, so the threshold can adapt to the local signal-to-noise ratio change of different frequency bands. This adaptive mechanism can effectively balance the contradiction between noise suppression and fault feature retention, which lays a high-quality data foundation for subsequent feature extraction and fault classification of the improved Transformer model.

3. Improved Transformer Model with Gated Multi-Head Attention

The standard Transformer architecture was originally designed to solve sequence modeling tasks within the field of natural language processing, and its core innovation lies in the multi-head self-attention mechanism. This mechanism breaks the sequential calculation restriction of recurrent neural networks, allowing the model to directly calculate the correlation degree between any two positions inside the entire input sequence. Such capability to capture long-range dependencies is particularly valuable for time-series vibration signal analysis, because fault-related impulse features may be separated by hundreds or thousands of sampling points in the one-dimensional vibration sequence. Traditional recurrent neural networks such as LSTM process signal samples step by step along the time axis, and the effective information of early sampling points will gradually decay after passing through multiple time steps, which makes it difficult to capture long-distance fault impulse correlations. Convolutional neural networks extract features only within fixed-size local receptive fields, and cannot actively establish connections between distant signal segments. The Transformer's attention mechanism overcomes both of these inherent limitations. When applied to vibration fault diagnosis, the one-dimensional vibration signal sequence is treated as a sequential sample, similar to a sentence composed of words in natural language tasks. Each sampling point of vibration signal is regarded as one basic element in the sequence, and the model learns the hidden feature representation of each sampling point through feature embedding.

Nevertheless, directly adopting the original vanilla Transformer for vibration signal fault classification still faces prominent practical defects. Even after adaptive wavelet denoising processing, the denoised vibration sequence still retains a certain amount of residual noise components and many redundant, non-informative signal segments. These segments do not contain fault impulse characteristics, and they belong to background signal components generated under normal equipment operating conditions. In the original multi-head self-attention module, attention weights are calculated indiscriminately for all sampling points in the entire sequence. The model will allocate equivalent computing resources and attention weight calculation to valid fault-related signal segments and irrelevant normal signal segments. This uniform calculation mode introduces interference from redundant features, increases the computational complexity of model training and inference, and diverts the model's attention away from the critical time segments that

© 2026 Javier Manuel Ortiz. This open-access manuscript is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits unrestricted reuse, distribution, and reproduction in any medium, provided the original author and source are fully credited. Full license text: https://creativecommons.org/licenses/by/4.0/

contain fault impact pulses. Especially under relatively high noise intensity, residual noise features will compete with fault features in the attention weight calculation process, which seriously weakens the model's classification performance for weak fault samples.

To mitigate this drawback, this paper embeds a learnable gated unit before the multi-head attention calculation module, and constructs a gated multi-head attention structure. The core function of the gated unit is to perform feature screening before attention weight calculation, so that the model can automatically judge the importance of different feature segments in the vibration sequence. The input of the gated unit is the high-dimensional feature embedding sequence obtained from the embedding layer. The gated unit first sends the embedded feature sequence into two independent fully connected layers. The first fully connected layer keeps the feature dimension unchanged and maps the original feature information; the second fully connected layer compresses the feature dimension and learns the importance distribution of each feature channel. The output of the second fully connected layer passes through a sigmoid activation function, which limits all output values to the interval between 0 and 1, forming a learnable gating weight matrix. Each element in this matrix corresponds to one feature channel of the vibration sequence. When the value of the gating weight is close to 1, it means that the corresponding feature channel contains high-value fault-related information and needs to be retained and enhanced; when the value approaches 0, it indicates that the feature channel mainly carries redundant or noise interference information, and the feature response of this channel will be suppressed.

After generating the gating weight matrix, element-wise multiplication is performed between the gating matrix and the original embedded feature sequence. This operation completes the adaptive filtering of the input features. The gated filtered feature sequence is then transmitted to the subsequent multi-head attention module for correlation calculation. The multi-head attention divides the feature space into multiple independent attention heads. Each attention head learns different types of feature association rules: some attention heads focus on the periodic impact characteristics of fault signals, while others learn the slow-changing trend characteristics of normal vibration signals. By combining the output results of multiple attention heads, the model can capture multi-scale correlation information inside the vibration sequence. Since the gated unit has already suppressed irrelevant feature channels before attention calculation, each attention head can concentrate its computing resources on the signal segments associated with fault features, instead of distributing computing power evenly across the entire signal sequence. This design greatly reduces the interference of useless residual features, improves the efficiency of feature learning, and strengthens the model's capability to identify weak fault impulses submerged in residual noise.

The overall workflow of the improved Transformer encoder is described in detail as follows. The denoised one-dimensional vibration signal sequence, which has been processed by adaptive wavelet denoising, is first sent into the embedding layer. The embedding layer converts the original low-dimensional sampling values of vibration signals into high-dimensional dense feature embedding vectors. This mapping operation can project simple vibration amplitude values into a higher-dimensional feature space, making it easier for the model to distinguish subtle differences between different fault types. Since the native Transformer structure does not have built-in perception of sequence position information, positional encoding is added to the embedded feature sequence. The positional encoding uses sine and cosine functions of different frequencies to inject the time position information of each sampling point. With positional encoding, the model can distinguish the time order of vibration sampling points, which is essential for analyzing periodic fault impact features.

After the completion of position encoding, the feature sequence enters the gated multi-head attention module. The feature sequence first passes through the gated unit to realize adaptive feature screening, and the gated filtered features are then used for multi-head self-attention calculation. The output of the gated multi-head attention module is added to the input feature sequence through residual connection, which can effectively alleviate the gradient vanishing problem during the deep model training process. After residual connection, layer normalization is performed on the feature data. Layer normalization standardizes the feature distribution inside each sample, stabilizing the distribution of intermediate features and accelerating the convergence speed of model training. The normalized feature sequence is then transmitted to the feed-forward network module. The feed-forward network consists of two fully connected layers with a ReLU activation function placed between them. This module performs non-linear transformation on each feature vector independently, further extracting and enriching the hidden feature information learned by the attention module. After the feed-forward network, another residual connection and layer normalization operation are executed, which completes the calculation process of one single encoder layer.

Multiple identical encoder layers are stacked sequentially to form the complete encoder part of the improved Transformer. With the stacking of encoder layers, the model can gradually extract high-level abstract fault features from the original vibration sequence. The shallow encoder layers mainly learn low-level features such as signal impact amplitude and periodic oscillation, while the deep encoder layers combine these low-level features to form

© 2026 Javier Manuel Ortiz. This open-access manuscript is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits unrestricted reuse, distribution, and reproduction in any medium, provided the original author and source are fully credited. Full license text: https://creativecommons.org/licenses/by/4.0/

high-level semantic features that can distinguish different fault categories. After the data passes through all stacked encoder layers, the global feature vector representing the whole vibration sample is obtained. This global feature vector is sent into the classification head at the end of the model. The classification head is composed of several fully connected layers and a softmax activation function. The fully connected layers map the global feature vector to the dimension corresponding to the number of fault categories, and the softmax function converts the output values into probability distribution among all fault categories. The category with the highest probability value is taken as the final diagnosis result, so as to realize the automatic classification of bearing fault types.

Compared with the original vanilla Transformer model, the gated multi-head attention module introduces an extra feature screening branch in the forward propagation process. In the model training phase, the parameters of the gated unit are optimized together with other parameters of the Transformer through backpropagation. The gate automatically learns the inherent distribution characteristics of fault features and normal background features from the training dataset. For feature segments that do not contain fault impulse information, the gate will automatically reduce their weight response; for feature segments containing periodic fault impact pulses, the gate will increase their feature weight and amplify their contribution to the final classification result. This adaptive feature filtering mechanism significantly improves the model's feature extraction efficiency, enhances the anti-interference performance under noisy working conditions, and makes the improved Transformer architecture more suitable for weak fault classification tasks of vibration time-series signals. In addition, the gated structure will not bring excessive computational overhead. Compared with the original Transformer, the growth of model parameter quantity is controlled within a reasonable range, so the model can maintain acceptable training and inference speed while obtaining obvious performance improvement.

4. Experimental Setup

4.1 Dataset Description

The rolling bearing dataset adopted in this paper is the widely used bearing fault dataset from Case Western Reserve University (CWRU). The vibration data is collected by acceleration sensors installed on the bearing housing. The dataset contains normal bearing state and three main fault types: inner race fault, outer race fault and rolling element fault. Each fault type includes multiple fault damage diameters, which can simulate different severity levels of bearing damage. In this experiment, vibration data under the same motor load condition is selected. The original long continuous vibration signal is segmented into multiple fixed-length one-dimensional sequence samples. Each sample contains a fixed number of sampling points. The dataset is randomly divided into training set, validation set and test set according to the proportion of seventy percent, fifteen percent and fifteen percent. The training set is used for model parameter optimization, the validation set is used for real-time monitoring of model overfitting during training, and the test set is used for final performance evaluation of the trained model. In order to verify the anti-noise performance of the model, different intensities of additional Gaussian white noise are added to the test samples in part of the comparison experiments, so as to simulate different low signal-to-noise ratio industrial working conditions.

4.2 Evaluation Metrics

Classification accuracy is selected as the main evaluation index of the model. Accuracy is defined as the ratio of the number of correctly diagnosed samples to the total number of test samples. In addition, F1-score is adopted as auxiliary evaluation metric, which is suitable for evaluating classification performance under the condition of unbalanced sample distribution. All models are trained repeatedly for multiple times, and the average value and standard deviation of the indexes are recorded to reduce the random fluctuation caused by model initialization.

4.3 Model Training Settings

The deep learning framework adopted in this experiment is PyTorch. The optimizer uses Adam optimization algorithm. The learning rate is set with a warm-up strategy at the initial stage of training, and the learning rate decays gradually with the increase of training epochs. The cross-entropy loss function is selected as the loss function for multi-class fault classification task. The batch size is set to a fixed value. The maximum training epoch is preset, and the early stopping strategy is adopted. When the loss value of the validation set no longer decreases continuously for multiple consecutive epochs, the training process will be terminated in advance to prevent model overfitting.

© 2026 Javier Manuel Ortiz. This open-access manuscript is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits unrestricted reuse, distribution, and reproduction in any medium, provided the original author and source are fully credited. Full license text: https://creativecommons.org/licenses/by/4.0/

5. Experimental Results and Discussion

Table 1 shows the classification accuracy comparison results of different models under the original test samples without additional noise. The comparison models include SVM, CNN, LSTM and vanilla Transformer. It can be observed from the data that the proposed integrated method combining adaptive wavelet denoising and gated multi-head attention Transformer achieves the highest average classification accuracy. Traditional machine learning method SVM relies on manually extracted features, and its performance is lower than deep learning models. CNN can extract local vibration features, but lacks the ability to capture long-range correlation in vibration sequences. LSTM can model time-series dependence, but long-distance information attenuation limits its performance. The original Transformer obtains better results than CNN and LSTM, but the lack of adaptive feature screening mechanism leads to lower accuracy than the improved model proposed in this paper.

Table 1

Classification Accuracy of Different Models under Noise-Free Condition

Model	Average Accuracy (%)	Standard Deviation
SVM	86.31	1.24
CNN	91.57	0.86
LSTM	92.14	0.93
Vanilla Transformer	94.62	0.71
Proposed Method	97.85	0.52

Note. APA 7 format table note: The table reports the average classification accuracy and standard deviation obtained after five repeated training runs for each model.

Table 2 presents the diagnostic accuracy of each model under test samples with added Gaussian white noise. With the increase of noise intensity, the classification accuracy of all models shows a declining trend. However, the proposed method has the slowest accuracy decline rate, which demonstrates that the combination of adaptive wavelet denoising and gated multi-head attention significantly improves the anti-noise robustness of the diagnosis framework. The adaptive wavelet denoising preprocessing removes most of the background noise, and the gated unit further suppresses residual noise features in the feature space, reducing the interference of noise components on fault feature learning. For weak fault samples, the model can still capture the faint periodic impact characteristics submerged in noise, which is the core advantage of this improved architecture.

Table 2

Classification Accuracy under Different Noise Intensity

Model	Low Noise (%)	Medium Noise (%)	High Noise (%)
SVM	81.22	72.45	63.17
CNN	87.63	79.32	68.74
LSTM	88.15	80.26	70.41
Vanilla Transformer	91.04	83.77	74.28
Proposed Method	95.36	90.18	84.63

Note. APA 7 format table note: Low, medium and high represent three levels of Gaussian white noise intensity added to the test set.

By analyzing the experimental results, we can find that the gated multi-head attention module brings two obvious improvements. On the one hand, the learnable gate completes feature screening before attention calculation, so the model can reduce the weight assigned to irrelevant background segments automatically, and focus computing resources on fault-related time segments. On the other hand, the gate is jointly optimized with the whole model during backpropagation, which realizes end-to-end adaptive feature selection without manually designing feature rules. Compared with the original Transformer,

© 2026 Javier Manuel Ortiz. This open-access manuscript is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits unrestricted reuse, distribution, and reproduction in any medium, provided the original author and source are fully credited. Full license text: https://creativecommons.org/licenses/by/4.0/

the improvement effect becomes more prominent under high noise conditions, which proves that the gated structure is targeted at solving the problem of weak fault feature recognition under noise interference.

It is also necessary to discuss the limitations of this research. The dataset used in this paper is collected under relatively stable motor load conditions. In actual industrial sites, the load of rotating machinery often changes continuously in real time. The generalization performance of the model under strong time-varying load needs further verification. In addition, the current research only carries out offline fault diagnosis based on collected vibration data. In future work, the model inference speed can be further optimized to adapt to the requirement of online real-time diagnosis on industrial equipment.

6. Conclusion

Aiming at the problem that weak fault impulse features of rolling bearings are easy to be submerged in industrial background noise, this paper constructs an intelligent fault diagnosis framework combining adaptive wavelet denoising and improved Transformer with gated multi-head attention. Adaptive wavelet threshold denoising is used as signal preprocessing module to suppress background noise while retaining transient fault impact features. A gated multi-head attention mechanism is embedded into Transformer encoder, which can automatically screen feature channels and suppress redundant noise features before attention calculation. The improved model can focus on the time segments containing fault impulses, enhance the mining ability of long-range fault correlation in vibration sequences, and improve the recognition performance of weak bearing faults under noisy conditions. Comparative experiments based on CWRU bearing dataset show that the proposed method achieves higher classification accuracy and stronger anti-noise robustness compared with SVM, CNN, LSTM and original Transformer.

In future research, we will further optimize the model structure to reduce model parameters and reasoning time, so as to adapt to the requirement of edge deployment on industrial field devices. Besides, multi-source sensor fusion technology will be introduced, combining vibration signals with temperature, current and other monitoring data, to further improve the reliability of rotating machinery fault diagnosis under complex variable working conditions.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest related to the publication of this paper.

Funding Statement

This research received no external funding.

Ethical Statement

This study only involves industrial vibration signal analysis and deep learning model simulation verification, and does not involve human participants or animal experiments. Therefore, ethical approval is not required for this study.

Data Availability Statement

The experimental data used in this study are partially derived from the public CWRU bearing fault dataset, which is openly available for academic research. The self-collected industrial gearbox weak fault test data are not publicly available due to industrial confidentiality restrictions. Qualified academic researchers can contact the corresponding author for reasonable data access applications for non-commercial research purposes.

7. References

Donoho, D. L. (1995). De-noising by soft-thresholding. IEEE Transactions on Information Theory, 41(3), 613–627.

Daubechies, I. (1992). Ten lectures on wavelets. Society for Industrial and Applied Mathematics.

Wang, Y., Xiang, J., Markert, R., & Liang, M. (2016). Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications. Mechanical Systems and Signal Processing, 66, 679–698.

© 2026 Javier Manuel Ortiz. This open-access manuscript is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits unrestricted reuse, distribution, and reproduction in any medium, provided the original author and source are fully credited. Full license text: https://creativecommons.org/licenses/by/4.0/

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213–237.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

Han, T., Liu, C., Yang, W., & Jiang, D. (2021). A novel adversarial learning framework for fault diagnosis of rolling bearings under varying working conditions. IEEE Transactions on Instrumentation and Measurement, 70, 1–10.

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, 205–217.

Zhang, X., Liu, Z., Wang, J., & Wang, H. (2022). Wavelet transform and improved Transformer for rolling bearing fault diagnosis under strong noise. Measurement, 191, 110815.

Chen, H., Jiang, B., Lu, N., & Mao, Z. (2020). Deep recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing, 409, 98–110.

Qiu, H., Lee, J., Lin, J., & Yu, H. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. Journal of Sound and Vibration, 289(4–5), 1066–1090.

Sahoo, S., & Jena, R. (2023). Adaptive wavelet threshold denoising for vibration signal of rotating machinery: A review. Journal of Vibration Engineering & Technologies, 11(2), 487–506.

Liu, Y., Yao, L., & Zhang, Y. (2022). Gated multi-head attention Transformer for time-series fault diagnosis of industrial machinery. ISA Transactions, 128, 541–551.

Ma, J., Wu, J., & Wang, X. (2020). A review of machine learning for vibration-based fault diagnosis of rotating machinery. Shock and Vibration, 2020, 1–21.

Yang, Y., & Qian, S. (2018). Wavelet denoising with adaptive threshold for mechanical vibration signals. Journal of Mechanical Science and Technology, 32(7), 3081–3088.

Hoang, D. T., & Kang, H. J. (2019). A survey on deep learning based bearing fault diagnosis. Neurocomputing, 335, 327–342.

Li, Y., Gao, L., & Wang, X. (2021). Weak fault feature extraction of rolling bearings based on improved wavelet threshold and spectral kurtosis. Journal of Sound and Vibration, 497, 115937.

Wang, Z., & Li, W. (2023). Anti-noise fault diagnosis of rotating machinery using wavelet denoising and Transformer with gated attention. Applied Acoustics, 204, 109246.

Rubinstein, R., & Donoho, D. L. (1999). Wavelet thresholding for unknown noise variance. Signal Processing, 74(3), 279–291.

Jiang, Q., Shen, Y., Li, H., & Zhang, L. (2020). Rolling bearing fault diagnosis using improved multi-scale wavelet transform and CNN. Measurement Science and Technology, 31(11), 115007.

Mousavi, A., & Bahrani, S. (2022). Kurtosis-based adaptive wavelet denoising for rotating machine vibration signals. Journal of Vibration and Control, 28(17–18), 2145–2158.

Zhang, H., Li, Y., & Chen, X. (2021). Time series fault classification based on Transformer with gating mechanism. Pattern Recognition, 116, 107948.

Wu, S., Prasad, M., & Wang, Y. (2022). Review of wavelet denoising techniques for machinery condition monitoring. Journal of Dynamic Systems, Measurement, and Control, 144(8), 081005.

Chen, Z., Li, W., & Yang, J. (2020). Vibration signal denoising using adaptive wavelet threshold optimized by genetic algorithm. IEEE Access, 8, 182116–182126.

Liu, J., & Wang, H. (2023). Robust fault diagnosis under heavy noise: Combining wavelet preprocessing and attention-based deep model. Journal of Manufacturing Systems, 68, 342–352.

Sharma, V., & Parey, A. (2020). A review of rolling bearing fault diagnosis techniques. Journal of Failure Analysis and Prevention, 20(2), 468–483.

© 2026 Javier Manuel Ortiz. This open-access manuscript is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits unrestricted reuse, distribution, and reproduction in any medium, provided the original author and source are fully credited. Full license text: https://creativecommons.org/licenses/by/4.0/

Li, M., & Zhao, J. (2022). Transformer-based deep learning for machinery fault diagnosis: Review and roadmap. Progress in Aerospace Sciences, 131, 100822.

Kedadouche, M., Thomas, M., & Tahan, A. (2016). A comparative study on empirical mode decomposition and wavelet transform for bearing fault diagnosis. Mechanical Systems and Signal Processing, 70, 254–273.

Wang, H., Chen, P., & Zhang, Y. (2021). Adaptive threshold wavelet denoising combined with LSTM for bearing fault diagnosis. Journal of Vibration Engineering, 34(4), 789–796.

Xu, Y., Yan, R., & Gao, R. X. (2022). Weak fault detection of bearings using wavelet packet transform and kurtosis-guided threshold. IEEE Transactions on Reliability, 71(3), 1432–1441.

Singh, J., & Kumar, A. (2023). Noise robust fault diagnosis of rotating machinery using wavelet denoising and modified Transformer network. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 45(5), 321.

© 2026 Javier Manuel Ortiz. This open-access manuscript is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license permits unrestricted reuse, distribution, and reproduction in any medium, provided the original author and source are fully credited. Full license text: https://creativecommons.org/licenses/by/4.0/