Journal of Theoretical and Applied Vibration and Acoustics 10(1) 54-66 (2024)

I S A V

Journal of Theoretical and Applied Vibration and Acoustics

journal homepage: http://tava.isav.ir

Noise-robust gearbox fault detection: A deep learning approach

Navidreza Ghanbari [a], Yasin Riyazi [a], Farzad A. Shirazi [b,*] Ahmad Kalhor [c]

[a] *M.Sc. Student, School of Mechanical Engineering, College of Engineering, University of Tehran, Tehran, IRAN*

[b] *Assistant Professor, School of Mechanical Engineering, College of Engineering, University of Tehran, Tehran, IRAN*

[c] *Associate Professor, School of Electrical Engineering, College of Engineering, University of Tehran, Tehran, IRAN*

Research Article

ARTICLE INFO

Article history:
Received 21 May 2024

Received in revised form
8 October 2024

Accepted 1 November 2024

Available online 6 January 2025

Keywords:
Gearbox fault diagnosis

Long Short-Term Memory (LSTM)

Convolutional Neural Networks (CNN)

Continuous Wavelet Transform (CWT)

Noise robustness

ABSTRACT

We introduce a novel approach to enhance gearbox fault diagnosis by integrating Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) for vibrational data analysis. Our method aims to improve fault detection accuracy, particularly in identifying subtle anomalies like broken teeth. However, real-world data often contains noise, which can hinder the effectiveness of such models. To address this challenge, we incorporate Singular Value Decomposition (SVD) pooling layers within the model. Our methodology starts with continuous wavelet transform (CWT), applied to the vibrational data to reveal crucial frequency-domain features. Concurrently, a CNN, using the Inception architecture, extracts spatial features. Simultaneously, LSTM networks capture temporal patterns. The unique feature representations from the CNN and LSTM branches are fused, creating a holistic feature set incorporating spatial, material, and frequency-domain information. This integrated feature set is then classified using a fully connected neural network. Our method's effectiveness is rigorously validated through comprehensive experiments on a diverse dataset. The results demonstrate exceptional accuracy in identifying gearbox faults, even in the early stages. This research advances predictive maintenance, offering a precise and comprehensive approach to gearbox fault diagnosis. In conclusion, the fusion of LSTM and CNN architectures for vibrational data analysis holds promise for gearbox fault diagnosis, benefiting industries reliant on machinery reliability and operational efficiency.

1. Introduction

Gearboxes play a critical role in machinery; their reliability is essential for uninterrupted industrial operations. Timely fault diagnosis is pivotal for preventing costly downtime and ensuring

* Corresponding author:
E-mail address: fshirazi@ut.ac.ir (F.A. Shirazi)

http://dx.doi.org/10.22064/tava.2024.2029883.1242

machinery longevity. Therefore, to guarantee safety, growing attention has been paid to fault diagnosis of gearboxes [1]. Previous methods aimed to develop a mathematical model to express specific faults, and some methods required prior knowledge for reasoning and diagnosis [2]. In modern problems, due to the complexity of engineering systems, developing a proper model is difficult [3]. Traditional machine learning algorithms have been widely used in the fault diagnosis field. Baraldi et al. [4] aimed to develop a diagnostic system for electric traction motor bearings in variable automotive conditions. Employing a hierarchical structure of K-Nearest Neighbors classifiers, this method selects relevant features from vibrational signals using a Multi-Objective optimization approach, showcasing its effectiveness across diverse operational conditions in experimental testing. These methods require manual feature extraction, relying heavily on human expertise.

Deep learning has grown rapidly in recent years, setting new performance standards. Chen et al. [5] used deep neural networks to effectively identify faults in rolling bearings, demonstrating their reliability in fault diagnosis. This is crucial for maintaining machinery performance and preventing mechanical failures. Jiang et al. [6] presented an end-to-end learning-based system that directly learns fault features from raw vibration signals. The method employs a multiscale convolutional neural network (MSCNN) that simultaneously extracts multiscale features, enhancing feature learning and diagnosis performance. Chen et al. [7] proposed an effective method utilizing convolutional neural networks (CNN) and discrete wavelet transformation (DWT) to diagnose fault conditions in planetary gearboxes used in wind turbines. Xie et al. [8] presented a gearbox fault diagnosis method using a PCNN–GRU model that fuses multi-sensor vibration data, achieving an accuracy rate of 99.92% in diagnosing faults. Guan et al. [9] proposed an improved transformer network for gearbox fault diagnosis, achieving 99.4% classification accuracy on five fault signals. Gao et al. [10] introduced an optimized adaptive deep belief network for rolling bearing fault diagnosis. The paper concludes with empirical validation through simulations based on experimental data, confirming the efficacy of the proposed method in bearing fault identification. Liang et al. [11] introduced WT-IResNet, a novel fault diagnosis method for rolling bearings based on wavelet transform and improved ResNet architecture. It addresses noisy labels and real-world industrial conditions through wavelet transform, an improved residual neural network, and a customized loss function. Xiao Et al. [12] proposed a novel fault diagnosis method for three-phase asynchronous motors using LSTM neural networks, which learn from raw data without feature engineering. Experimental tests are more accurate than traditional methods like LR, SVM, MLP, and RNN.

This study presents a novel and robust deep-learning approach for gearbox fault diagnosis. We leverage the strengths of Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) for comprehensive analysis of vibrational data. This fusion captures spatial and temporal features crucial for identifying gearbox faults, particularly subtle anomalies like broken teeth. Our methodology incorporates Singular Value Decomposition (SVD) pooling layers within the CNN architecture to address the challenge of noise in real-world data. SVD pooling enhances the model's noise robustness, leading to superior performance even with noisy data.

Building upon the strengths of CWT for frequency-domain feature extraction and the Inception architecture for spatial features, we employ LSTM networks to capture temporal dependencies. The resulting features are then processed for classification. This integration of LSTM and CNN, along with feature fusion and noise-robust SVD pooling, holds significant promise for accurate

gearbox fault diagnosis. This research advances predictive maintenance by enabling early fault detection and enhancing machinery reliability and operational efficiency. In the following sections, we delve into the theoretical background, detail our methodology, present the experimental findings, and discuss the results, evaluating the effectiveness of the proposed approach for robust gearbox fault diagnosis.

2. Theoretical foundation

2.1. Continuous Wavelet Transform

CWT [13] is a mathematical technique that simultaneously analyzes signals in both the time and frequency domains. It provides a way to examine how the frequency content of a signal evolves. This is particularly useful when dealing with non-stationary signals, whose characteristics change over different time intervals. CWT can effectively decompose the initial signal into various oscillatory components, which originate from the translation and scaling of mother wavelets [14].

The CWT of a signal f(t) is calculated as shown in Equation (1):

$$WT(a.\tau) = \int_{-\infty}^{\infty} f(t) \cdot \psi^* \left(\frac{t-\tau}{a}\right) dt, \tag{1}$$

where $f(t)$ is the input signal, $\psi(t)$ is the mother wavelet, ψ^* is the complex conjugate of the mother wavelet, τ is the translation parameter, which shifts the wavelet function along the time axis to analyze different time points in the signal, and a is the scale parameter, which controls the width of the wavelet function and determines the level of detail in the analysis.

2.2. Convolutional Neural Network

CNN [15] is a class of deep-learning neural networks primarily designed for processing structured grid data, such as images and video. It is inspired by the human visual system and is highly effective in tasks like image classification, object detection, and image segmentation. CNNs begin with one or more convolutional layers. These layers apply filters (also known as kernels) to the input image. Each filter is a small matrix that scans through the input using a mathematical operation called convolution. The convolution operation extracts features like edges, textures, or patterns from the input. Figure 1 shows the schematic of a CNN containing convolution, pooling, and fully connected layers.

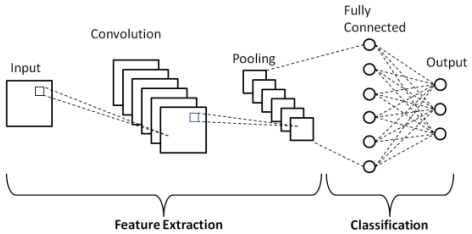

Fig. 1. Schematic of a CNN [16].

2.2. Inception module

The Inception architecture, a seminal advancement in deep CNNs, represents a pivotal approach in neural network design, notable for its unparalleled capacity to capture intricate spatial features from multidimensional data. Introduced by Szegedy et al. [17], Inception tackles the challenge of effective feature extraction and dimensionality reduction. At its core, Inception employs multiple filter sizes and operations within a single layer. Unlike conventional layers with fixed-sized filters, Inception simultaneously uses various filter sizes to capture information at different spatial scales. This multiscale approach helps in capturing both fine and coarse spatial details. In addition, Inception incorporates dimensionality reduction techniques like 1x1 convolutions and pooling to reduce computational complexity while preserving essential features. Integrating the Inception architecture into the CNN branch enhances the model's ability to discern critical information efficiently and accurately. The schematic of the Inception model is shown in Figure 2.

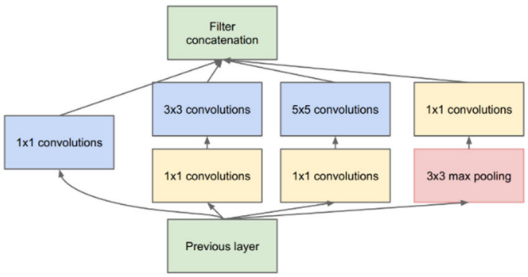

Fig. 2. Inception Module.

2.3. Long Short-Term Memory

A Long Short-Term Memory (LSTM) [18] is a type of Recurrent Neural Network (RNN) architecture designed to handle sequential data and address the vanishing gradient problem that traditional RNNs often face. LSTMs are designed for tasks involving data sequences, such as time series, natural language text, or speech. LSTMs maintain a cell state, which serves as a memory buffer. This cell state can carry information across time steps and selectively forget or update information, making it well-suited for capturing long-range dependencies in sequences. In addition to the cell state, LSTMs also maintain a hidden state. This hidden state serves as the memory that carries information to the next time step.

2.4. SVD pooling

SVD is a mathematical technique for decomposing a real or complex matrix into a set of constituent components. Given a matrix $\mathbf{X}$ of size (m×n), SVD decomposes it into three matrices:

- $\mathbf{U}$ (m×m): An orthogonal matrix containing the left singular vectors of $\mathbf{X}$,
- $\mathbf{\Sigma}$ (m×n): A diagonal matrix containing the singular values of $\mathbf{X}$ in non-increasing order,
- $\mathbf{V}$ (n×n): An orthogonal matrix containing the right singular vectors of $\mathbf{X}$.

The decomposition can be expressed as:

$$\mathbf{X} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{T}$$

The singular values in **Σ** represent the importance of the corresponding singular vectors in **U** and **V**. The first few singular values capture the most significant information within the matrix, while the later ones represent progressively smaller variations.

Common pooling layers in CNNs, such as max pooling and average pooling, operate by applying a pre-defined window across the feature map and aggregating the values within that window. While these methods effectively reduce dimensionality, they have limitations when dealing with noisy data. Max pooling can be overly sensitive to outliers, potentially discarding valuable information due to a single noisy spike in a window. On the other hand, average pooling might be susceptible to the overall noise level within the window, potentially amplifying noise if most values are corrupted.

In contrast, SVD pooling offers a more nuanced approach. By decomposing the feature map into its constituent singular vectors and corresponding singular values, SVD pooling retains the most significant information (represented by the larger singular values) while discarding noise-induced variations reflected in the smaller singular values. This selective retention process allows SVD pooling to capture the underlying structure of the data and suppress noise, leading to a more robust feature representation. This robustness is particularly beneficial for real-world scenarios like gearbox fault diagnosis, where the data may be corrupted by noise from various sources during data acquisition.

3. Proposed method

3.1. Framework of the proposed method

Gearboxes are susceptible to faults due to their exposure to various operational stresses and environmental conditions. Effective fault diagnosis allows for the early detection and mitigation of these issues, preventing costly downtime and reducing maintenance expenses. This process is fundamental to ensuring machinery reliability and the uninterrupted flow of industrial operations.

To prevent the above problems, in this study, we propose a new model for fault diagnosis of the gearbox, named the fusion CNN-LSTM model. This model consists of three main parts: the CNN model, the LSTM model, and classification layers. Figure 3 reveals the schematic of the model. The detailed steps are as follows:

1. Raw vibrational data are fed to an LSTM, analyzing the temporal dynamics within the vibrational data. LSTM is capable of capturing sequential dependencies and nuanced variations over time.
2. In parallel with the LSTM, through CWT, original data are converted into images and fed to a CNN, extracting spatial features from the data and identifying distinctive patterns and spatial relationships that can aid in fault diagnosis.
3. The outputs from the CNN and LSTM branches are combined. This feature stacking creates a comprehensive representation of the vibrational data, incorporating both spatial and temporal information.
4. The integrated feature set is passed to a fully connected neural network for the classification task. The neural network determines whether the gearbox is operating normally or experiencing a fault based on the combined feature representation.

5. The model's performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques are employed to assess the model's robustness and generalizability.

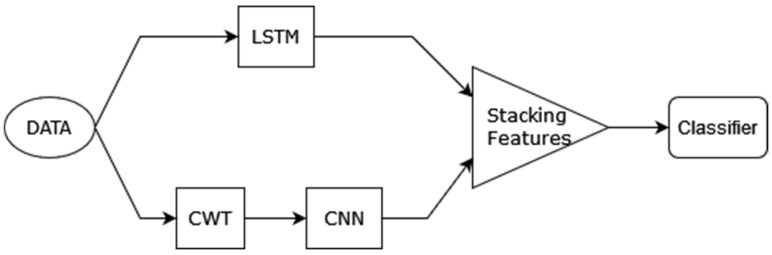

Fig. 3. The overall framework.

3.2. LSTM branch

The LSTM branch begins with the input of vibrational data, a time-series sequence collected from the gearbox. This data is crucial for capturing the temporal dynamics of the system. Before feeding the data into the LSTM network, pre-processing steps such as normalization and sequence length adjustment are applied to ensure data consistency and compatibility with the network. The core of the LSTM branch consists of one LSTM layer. As the data flows through the LSTM layer, the network analyzes the sequential patterns and dependencies within the vibrational data. LSTM units have the unique ability to capture both short-term and long-term temporal dependencies, making them well-suited for time-series data like vibration signals.

The LSTM branch produces either a sequence of hidden states or a summarization of the sequential analysis. Subsequently, an MLP layer is employed to project these hidden states into the desired dimensional space. The LSTM layer configuration remains consistent throughout the study, comprising a single hidden layer with 50 hidden states and yielding 60 output features. These specific parameter values were determined through an extensive grid hyperparameter search process.

3.3. CNN branch

First, continuous wavelet transform is applied to the original vibrational data to reveal frequency-domain features. CWT enables the extraction of intricate temporal patterns, enhancing the model's ability to identify gearbox faults accurately by capturing subtle variations in the data. Wavelet functions in CWT serve as analysis tools, each representing a specific frequency and time domain. The choice of wavelet function impacts the scale at which features are detected.

The Inception architecture is known for efficiently extracting spatial features from complex data. Therefore, it is used to extract meaningful features from vibrational data. The proposed method is shown in Figure 4. Table 1 demonstrates the network details. After concatenating filters, Global Average Pooling (GAP) is applied to reduce dimensions. GAP acts as a form of spatial information summarization, producing a compact representation that retains essential features while significantly reducing computational complexity. This operation is particularly beneficial for model efficiency, regularization, and interpretability, making it a fundamental component in various computer vision tasks.

In the initial design of our model, GAP served the purpose of dimensionality reduction, transforming the feature maps into a lower-dimensional vector suitable for classification. However, GAP's inherent characteristic of averaging across all spatial locations can be susceptible to noise in the data. To address this limitation and enhance the model's robustness, we incorporated SVD pooling. SVD pooling offers the dual benefit of dimensionality reduction while mitigating the impact of noise. By decomposing the feature maps and retaining the dominant singular components, SVD pooling captures the underlying structure of the data and discards noise-induced variations. This approach leads to a more robust feature representation, ultimately improving the model's ability to achieve accurate fault classification even in the presence of noisy gearbox vibration data.

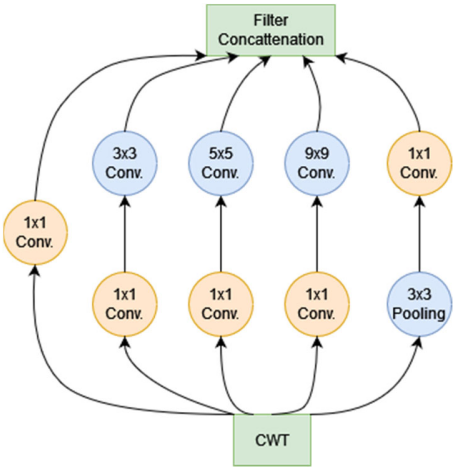

Fig. 4. Proposed inception module.

Table 1. Parameters in the CNN

	Seq. Block	input	output	kernel size	padding	stride
Branch1x1	Conv. 1×1	1	10	1	0	1
	ReLU	-	-	-	-	-
Branch3x3	Conv. 1×1	1	10	1	0	1
	ReLU	-	-	-	-	-
	Conv. 3×3	10	10	7	'same'	1
	ReLU	-	-	-	-	-
Branch5x5	Conv. 1×1	1	10	1	0	1
	ReLU	-	-	-	-	-
	Conv. 5x5	10	10	5	'same'	1
	ReLU	-	-	-	-	-
Branch9x9	Conv. 1×1	1	10	1	0	1
	ReLU	-	-	-	-	-
	Conv. 9×9	10	10	9	'same'	1
	ReLU	-	-	-	-	-
Pooling Branch	MaxPool2d	1	1	1	1	1
	-	-	-	-	-	-
	Conv. 1×1	1	10	1	0	1
	ReLU	-	-	-	-	-

3.4. Training and optimization

After stacking features are extracted by CNN and LSTM branches, GAP is applied to reduce dimensions. The integrated feature representation is fed into a Fully Connected Neural Network (FCNN). This neural network is responsible for the final classification task, distinguishing between healthy and faulty gearbox conditions. The FCNN's architecture typically consists of multiple layers of neurons, allowing it to learn complex relationships within the combined feature set. The proposed model is trained on Gearbox Fault Diagnosis Data [19] collected in the National

Renewable Energy Laboratory (NREL). This dataset includes examples of healthy and broken tooth gearbox conditions recorded under load variation from '0' to '90' percent load, providing useful samples for training. The health condition of the gear has a remarkable impact on the vibrational characteristic of the gearbox. Figure 5 shows a gear with broken teeth. During training, the network adjusts its internal parameters (weights and biases) through backpropagation and gradient descent to minimize the classification error. This phase is crucial for the model to classify the data accurately.

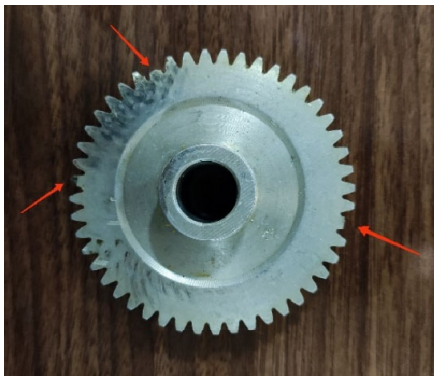

Fig. 5. Gear with broken teeth.

To evaluate the model's robustness against noise, our training process involved modifying the base model architecture to incorporate SVD pooling for enhanced noise robustness. We replaced the standard GAP layer with an SVD pooling layer. This substitution aims to capture the underlying structure of the feature maps while mitigating the impact of noise. Following the SVD pooling layer, we added two fully connected layers to learn a non-linear mapping between the extracted features and a lower dimension. The entire network was then trained using the Adam optimizer and a suitable cross-entropy loss function to minimize the classification error on the training data.

3.5. Performance metrics

The classification results are measured using performance metrics such as accuracy and F1-score. Accuracy measures the proportion of correctly classified instances in a dataset, expressed as a percentage. It indicates how often the model's predictions are correct overall. The expression of accuracy is shown in Equation (2).

$$acc = \frac{TP + TN}{TP + TN + FN + FP}, \tag{2}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives.

The F1-score is a single value that balances precision (accuracy of positive predictions) and recall (ability to find all relevant positive instances). It provides a comprehensive performance measure. The formula of the F1-score is given in Equation (3).

$$\mathrm{F1-score} = \frac{2 \cdot (\mathrm{Precision} \cdot \mathrm{Recall})}{\mathrm{Precision} + \mathrm{Recall}}, \tag{3}$$

where Precision is the ratio of true positive predictions to the total number of positive predictions made by the model, and Recall is the ratio of true positive predictions to the total number of actual positive instances in the dataset.

4. Results

4.1. Performance without noise in data

In this paper, different window sizes for CWT have been tested, namely 17, 50, and 100. The results are presented below in Tables 2 and 3. Based on the findings, the window size of 50 yielded the most favorable outcomes in the CNN-LSTM model. It is noticeable that the separate CNN model has the best performance with a window size of 17. However, when combined with LSTM, our proposed method yields the best performance with a window size of 50. The data in these experiments were obtained from a noise-free dataset recorded using a gearbox fault diagnostics simulator [19], ensuring the results reflect the model's performance in an ideal, controlled environment.

Table 2. CNN Model Results

Window size	F1-score	Support
17	0.98	880
50	0.94	880
100	0.35	880

Table 3. CNN-LSTM Model Results

Window size	F1-score	Support
17	0.98	880
50	1	880
100	0.41	880

.

The remaining model parameters were fine-tuned using a small grid hyperparameter search method [20], [21]. Further details regarding the specific hyperparameters and their ranges can be found in the referenced sources. The final diagnosis results of the model are presented in Table 4. It can be seen that this method has a remarkable performance.

Table 4. Final results for window size=50

	Precision	Recall	F1-score	Support
Faulty	1	1	1	880
Healthy	1	1	1	880
Accuracy	-	-	1	1760
Macro avg.	1	1	1	1760
Weighted avg.	1	1	1	1760

To demonstrate the enhancement of this method, it is compared with each of its branches separately as a model. It can be seen that the f1-score of the CNN model and LSTM model is 0.94 and 0.98, respectively. However, as this paper proposes, it would rise to 1 when their features are stacked together. This comparison is shown in Table 5.

Table 5. Comparison of the proposed model

Model	F1-score	Support

CNN	0.94	880
LSTM	0.98	880
CNN-LSTM	1	880

4.2. Model evaluation with noisy data

To evaluate the model's robustness against noise, we introduced controlled noise to the training and testing datasets. Here, 'controlled noise' refers to systematically added noise with pre-defined levels, allowing for consistent and replicable testing conditions. The noise was simulated to mimic real-world scenarios encountered in gearbox vibration data acquisition (e.g., sensor noise and environmental factors). We experimented with different noise levels to assess the model's ability to maintain performance under varying noise conditions.

We employed a rigorous testing procedure to assess noise's impact on model performance. The test data was augmented with varying noise amplitudes, simulating real-world scenarios encountered during data acquisition (e.g., sensor noise and environmental factors). The base model with GAP pooling and the model incorporating SVD pooling were evaluated on each noise level. The testing process was repeated 50 times for each noise level to ensure robustness and reduce the influence of outliers. The reported performance metrics represent 95% of the obtained accuracy values across these repetitions. This approach provides a statistically robust evaluation of the models' performance under varying noise conditions. The results and corresponding error bounds are depicted in Figure 6, where the y-axis represents the model's accuracy, and the x-axis represents the ratio of signal variance to noise variance (SNR) calculated in a logarithmic scale using Equation 4.

$$10\log\left(\frac{W_s}{W_n}\right) = 20\log\left(\frac{\sigma_s}{\sigma_n}\right) \tag{4}$$

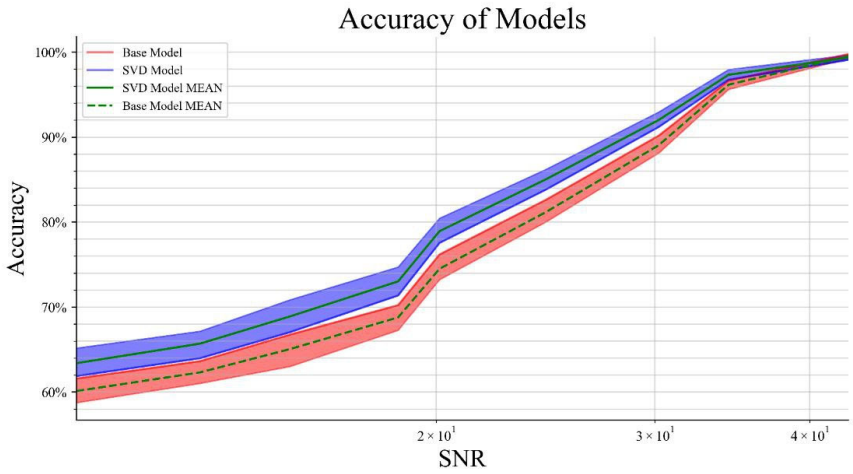

Fig.6. Performance of both models with different signal-to-noise ratios.

The impact of noise on model performance is shown in Figure 6. As expected, both the base and SVD models exhibit high accuracy when noise is low (higher SNRs). However, the performance

of both models degrades with increasing noise levels. Notably, the SVD model demonstrates superior robustness against noise compared to the base model. This is evident from the consistently higher accuracy of the SVD model across various noise levels in Figure 6.

Figure 7 quantifies this observation by presenting the difference in mean accuracy between the two models. The widening gap between the lines signifies the increasing advantage of the SVD model as noise intensifies (decreasing SNRs). However, both models experience significant performance drops at very high noise levels. This is likely because the data becomes heavily corrupted, making distinguishing noise from the underlying signal challenging and leading to classification errors.

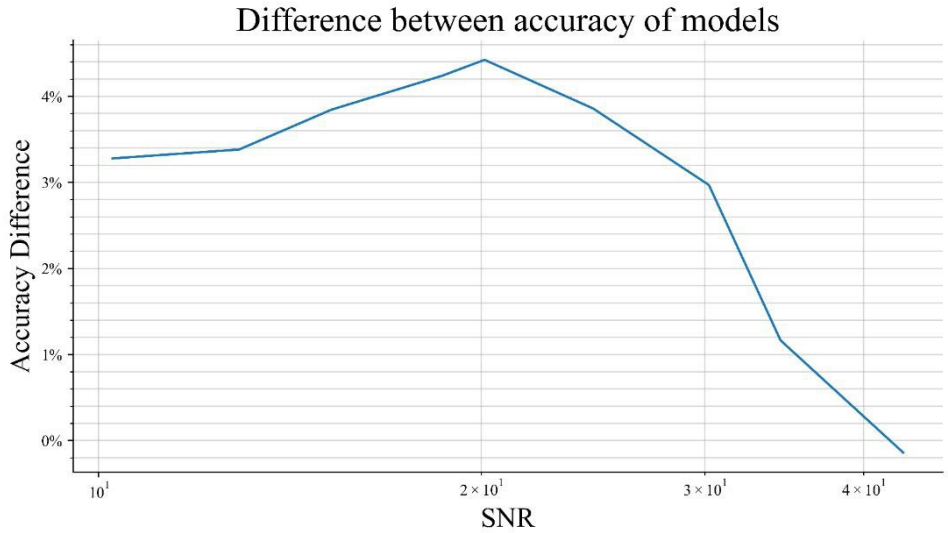

Fig. 7. Difference between the accuracy of models.

In conclusion, these results demonstrate the effectiveness of SVD pooling in enhancing the model's ability to handle noise. The SVD model consistently outperforms the base model under realistic noisy conditions (commonly encountered in industrial settings), highlighting its potential for robust gearbox fault diagnosis.

5. Conclusions

This study presents an innovative approach for enhancing gearbox fault diagnosis by integrating LSTM networks and CNNs. Our research aimed to leverage the strengths of these architectures to provide a comprehensive and accurate analysis of vibrational data, ultimately advancing the state-of-the-art in machinery condition monitoring. By integrating LSTM and CNN, we achieved exceptional accuracy in identifying gearbox faults, even in their early development stages. The fusion of spatial and temporal insights provided by these two architectures created a holistic feature representation that enhanced our model's fault detection capabilities. Our methodology, which included CWT for frequency-domain feature extraction and the Inception architecture for spatial feature extraction, showcased its robustness and generalizability through rigorous evaluation and cross-validation. We demonstrated the model's effectiveness in real-world scenarios, where early fault detection proved instrumental in reducing downtime and operational costs.

Furthermore, we incorporated SVD pooling layers to address noise challenges in real-world data. This approach significantly improves the model's noise robustness, as demonstrated by its superior performance under varying noise levels compared to a baseline model. This robustness is crucial for practical applications where sensor noise and environmental factors can corrupt data. Our work contributed to advancing predictive maintenance strategies through a robust and efficient deep learning approach for improved efficiency, reliability, and sustainability in industrial operations.

References

[1] D. Zhao, T. Wang, F. Chu, Deep convolutional neural network based planet bearing fault classification, Computers in Industry, 107 (2019) 59-66.

[2] Q. Guo, X. Zhang, J. Li, G. Li, Fault diagnosis of modular multilevel converter based on adaptive chirp mode decomposition and temporal convolutional network, Engineering Applications of Artificial Intelligence, 107 (2022) 104544.

[3] J. Xu, L. Zhou, W. Zhao, Y. Fan, X. Ding, X. Yuan, Zero-shot learning for compound fault diagnosis of bearings, Expert Systems with Applications, 190 (2022) 116197.

[4] P. Baraldi, F. Cannarile, F. Di Maio, E. Zio, Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions, Engineering Applications of Artificial Intelligence, 56 (2016) 1-13.

[5] Z. Chen, X. Chen, C. Li, R.-V. Sanchez, H. Qin, Vibration-based gearbox fault diagnosis using deep neural networks, Journal of Vibroengineering, 19 (2017) 2475-2496.

[6] G. Jiang, H. He, J. Yan, P. Xie, Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox, IEEE Transactions on Industrial Electronics, 66 (2018) 3196-3207.

[7] R. Chen, X. Huang, L. Yang, X. Xu, X. Zhang, Y. Zhang, Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform, Computers in industry, 106 (2019) 48-59.

[8] F. Xie, G. Wang, J. Shang, E. Sun, S. Xie, Gearbox Fault Diagnosis Based on Multi-Sensor Deep Spatiotemporal Feature Representation, Mathematics, 11 (2023) 2679.

[9] K. Guan, B. Du, Z. Wu, J. Li, Y. Zhang, Fault Diagnosis of Gearbox Based on Improved Transformer, in: Proceedings of the 2023 International Conference on Advances in Artificial Intelligence and Applications, 2023, pp. 312-318.

[10] S. Gao, L. Xu, Y. Zhang, Z. Pei, Rolling bearing fault diagnosis based on SSA optimized self-adaptive DBN, ISA transactions, 128 (2022) 485-502.

[11] P. Liang, W. Wang, X. Yuan, S. Liu, L. Zhang, Y. Cheng, Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment, Engineering Applications of Artificial Intelligence, 115 (2022) 105269.

[12] D. Xiao, Y. Huang, X. Zhang, H. Shi, C. Liu, Y. Li, Fault diagnosis of asynchronous motors based on LSTM neural network, in: 2018 prognostics and system health management conference (PHM-Chongqing), IEEE, 2018, pp. 540-545.

[13] P. Rastogi, E. Hack, Phase estimation in optical interferometry, CRC Press, 2014.

[14] D. Wang, Y. Zhao, C. Yi, K.-L. Tsui, J. Lin, Sparsity guided empirical wavelet transform for fault diagnosis of rolling element bearings, Mechanical systems and signal processing, 101 (2018) 292-308.

[15] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86 (1998) 2278-2324.

[16] V.H. Phung, E.J. Rhee, A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets, Applied Sciences, 9 (2019) 4500.

[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer

vision and pattern recognition, 2015, pp. 1-9.
[18] S. Hochreiter, Long Short-term Memory, Neural Computation MIT-Press, (1997).
[19] Y. Pandya, Gearbox fault diagnosis data, OpenEI National Renewable Energy Laboratory, USA, (2018).
[20] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics), 2009.
[21] C. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), 2007.