Enhanced Gearbox Fault Diagnosis with Fusion LSTM-CNN Network

NavidReza Ghanbaria, Yasin Riyazia, Farzad A. Shirazib

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.

\* Corresponding author e-mail: [fshirazi@ut.ac.ir](mailto:fshirazi@ut.ac.ir) (F. A. Shirazi)

**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. 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 both the CNN and LSTM branches are fused, creating a holistic feature set incorporating spatial, temporal, 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. The model's ability to detect faults promptly empowers industrial operators to reduce downtime and operational costs. 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.

**Keywords**: Gearbox Fault Diagnosis; Long Short-Term Memory (LSTM); Convolutional Neural Networks (CNN); Continuous Wavelet Transform (CWT).

# Introduction

Gearboxes play a critical role in machinery, and their reliability is essential for uninterrupted industrial operations. Timely fault diagnosis is pivotal for preventing costly downtime and ensuring machinery longevity. Therefore, to guarantee safety, growing attention has been payed to fault diagnosis of gearboxes[1]. In previous methods, the aim was 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 complexity of engineering systems, developing a proper model is difficult[3]. Traditional machine learning algorithm have been widely used in 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.

In recent years, deep learning has grown rapidly, 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, which is crucial for maintaining machinery performance and preventing mechanical failures. Jiang et al.[6] presents 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. Gao et al.[8] 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.[9] introduced WT-IResNet, a novel fault diagnosis method for rolling bearings, based on wavelet transform and improved ResNet architecture. It effectively addresses noisy labels and real-world industrial conditions through wavelet transform, an improved residual neural network, and a customized loss function. Xiao Et al.[10] 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 demonstrate superior accuracy compared to traditional methods like LR, SVM, MLP, and RNN.

This study presents an innovative approach to gearbox fault diagnosis, combining LSTM and CNN. Vibrational data collected under both healthy and faulty conditions is analyzed, focusing on identifying broken teeth as a common fault scenario. This method begins with Continuous Wavelet Transform (CWT) applied to the data, which is then processed by CNNs to extract spatial features. Simultaneously, LSTM networks capture temporal dependencies within the data. The resulting features from both networks are stacked and used for classification. This integration of LSTM and CNN, along with feature fusion, holds promise for accurate gearbox fault diagnosis. This research contributes to predictive maintenance, enhancing machinery reliability. In the following sections, we present our methodology, experimental findings, and discussions, evaluating the approach's effectiveness in gearbox fault diagnosis.

# Theoretical Foundation

## Continuous Wavelet Transform:

The Continuous Wavelet Transform (CWT)[11] is a mathematical technique employed to analyse signals in both the time and frequency domains simultaneously. It provides a way to examine how the frequency content of a signal evolves over time. This is particularly useful when dealing with non-stationary signals, where the signal's characteristics change over different time intervals. Over the past decade, continuous wavelet transform has seen significant advancements as a time-frequency analysis technique. CWT can effectively decompose the initial signal into various oscillatory components, which originate from the translation and scaling of mother wavelets[12].

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

where is the input signal, is the mother wavelet, is the complex conjugate of the mother wavelet, is the translation parameter, which shifts the wavelet function along the time axis to analyse different time points in the signal, and is the scale parameter, which controls the width of the wavelet function and determines the level of detail in the analysis. Smaller values of correspond to analysing high-frequency components, while larger values of correspond to low-frequency components.

## Convolutional Neural Network

Convolutional Neural Network (CNN)[13] is a class of deep learning neural networks primarily designed for processing structured grid data, such as images and video. It's 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. CNNs are trained using backpropagation and gradient descent. They learn to adjust the parameters (weights and biases) of the filters and fully connected layers to minimize a loss function, effectively fine-tuning themselves to recognize patterns in the training data. CNN has revolutionized computer vision tasks and is widely used in various applications, from image classification to object detection and beyond. Fig. 1 shows schematic of a CNN, containing convolution, pooling, and fully connected layers.



Figure 1. Schematic of a CNN [14].

### Inception Module

The Inception architecture, a seminal advancement in deep convolutional neural networks (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.[15], 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 multi-scale 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. This adaptability and feature diversity make Inception a powerful choice for tasks like extracting spatial patterns from vibrational data for gearbox fault diagnosis. Integrating the Inception architecture into the CNN branch enhances the model's ability to discern critical information efficiently and accurately. Schematic of inception model is shown on Fig. 4.

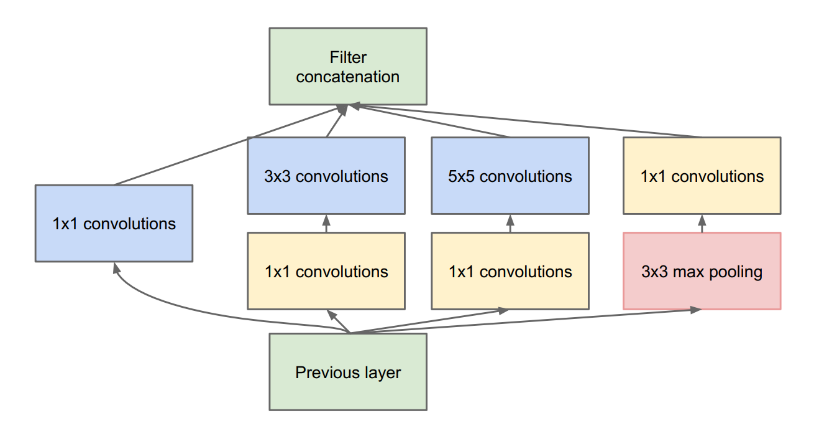


Figure 2. Inception Module.

## Long Short-Term Memory

A Long Short-Term Memory (LSTM)[16] 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 specifically designed for tasks involving sequences of data, such as time series, natural language text, or speech. They excel at capturing dependencies and patterns over time, making them suitable for tasks like speech recognition, machine translation, and sentiment analysis. LSTMs maintain a hidden state known as the 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. LSTMs are trained using backpropagation through time (BPTT). They learn to adjust the parameters (weights and biases) of their gates to minimize a loss function, effectively learning how to process and remember information across sequential data. Architecture of LSTM is shown below in Fig. 2.

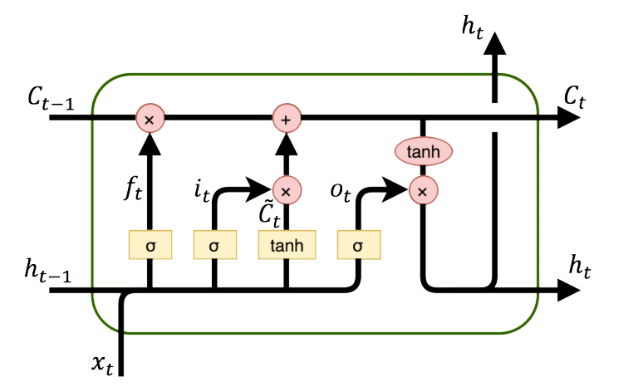


Figure 3. LSTM architecture.

# Proposed Method

## Framework of proposed method

Due to their exposure to various operational stresses and environmental conditions, gearboxes are susceptible to faults. Effective fault diagnosis allows for the early detection and mitigation of these issues, preventing costly downtime, reducing maintenance expenses, and prolonging the operational life of machinery. 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 gearbox, named fusion CNN-LSTM model. This model consists of three main parts, CNN model, LSTM model, and classification layers. Fig. 3 reveals the schematic of model. The detailed steps are as follows:

1. Raw vibrational data are fed to an LSTM, analysing 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 wavelet transform, original data are converted into images and fed to a CNN, extracting spatial features from the data, identifying distinctive patterns and spatial relationships that can aid in fault diagnosis.
3. Combine the outputs from the CNN and LSTM branches. This feature stacking creates a comprehensive representation of the vibrational data, incorporating both spatial and temporal information.
4. Pass the integrated feature set 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 amalgamated feature representation.
5. Rigorously evaluate the model's performance using standard metrics such as accuracy, precision, recall, and F1-score. Employ cross-validation techniques to assess the model's robustness and generalizability.

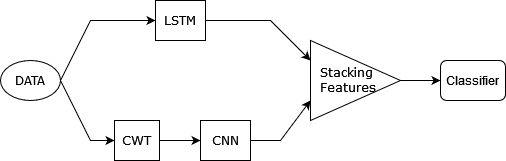


Figure 4. The overall framework.

## LSTM Branch

The LSTM branch begins with the input of vibrational data, which is 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, preprocessing 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 [TBD] LSTM layers. As the data flows through the LSTM layer(s), 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 output of the LSTM branch typically consists of a sequence of hidden states or a summary of the sequential analysis.

## CNN Branch

First, continuous wavelet transform is applied to 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 celebrated for its ability to efficiently extract spatial features from complex data. Therefore, it is used to extract meaningful features from vibrational data. Proposed method is shown in Fig. 5. Table 1 demonstrates details of network. 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.

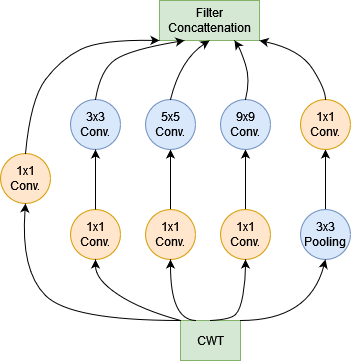
 Table 1. Parameters on CNN

Figure 5. Proposed inception module.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 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 | - | - | - | - | - |

## Training and Optimization

After stacking features extracted by CNN and LSTM branch, GPA 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 FCNN is trained on labeled data, which includes examples of both healthy and faulty gearbox conditions. 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 learn to accurately classify the data.

## 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 Eq. (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 F1-score is as follows:

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.

# Results

During this paper, different window sizes for Continuous Wavelet Transform (CWT) have been tested, namely 17, 50, and 100. The results are presented in Table 2 and Table 3 below Based on the findings, the window size of 50 yielded the most favorable outcomes in the CNN-RNN model.

Table 2. CNN Model Results

|  |  |  |
| --- | --- | --- |
| window size | f1-score | support |
| 17 | 0.99 | 880 |
| 50 | 0.94 | 880 |
| 100 | 0.35 | 880 |

Table 3. CNN-RNN 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 [insert reference]. Further details regarding the specific hyperparameters and their ranges can be found in the referenced source.

# Conclusions

In this study, we have presented 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. Through the integration of LSTM and CNN, we achieved exceptional accuracy in identifying gearbox faults, even in the early stages of their development. 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 Continuous Wavelet Transform (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.

As we look to the future, further exploration of hybrid deep learning approaches, like the one presented here, holds promise in addressing complex industrial challenges. We anticipate that our research will inspire continued innovation in predictive maintenance strategies and foster collaboration between the fields of machine learning and industrial engineering. Ultimately, this work contributes to the goal of achieving greater efficiency, reliability, and sustainability in industrial operations.

REFERENCES

1. Zhao, D., T. Wang, and F. Chu, *Deep convolutional neural network based planet bearing fault classification.* Computers in Industry, 2019. **107**: p. 59-66.

2. Guo, Q., et al., *Fault diagnosis of modular multilevel converter based on adaptive chirp mode decomposition and temporal convolutional network.* Engineering Applications of Artificial Intelligence, 2022. **107**.

3. Xu, J., Zhou, et al., *Zero-shot learning for compound fault diagnosis of bearings.* Expert Systems with Applications, 2022. **190**.

4. Baraldi, P., et al., *Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions.* Engineering Applications of Artificial Intelligence, 2016. **56**: p. 1-13.

5. Chen, Z., et al., *Vibration-based gearbox fault diagnosis using deep neural networks.* Journal of Vibroengineering, 2017. **19**(4): p. 2475-2496.

6. Jiang, G., et al., *Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox.* IEEE Transactions on Industrial Electronics, 2019. **66**(4): p. 3196-3207.

7. Chen, R., et al., *Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform.* Computers in Industry, 2019. **106**: p. 48-59.

8. Gao, S., et al., *Rolling bearing fault diagnosis based on SSA optimized self-adaptive DBN.* ISA Transactions, 2021. **128**: p. 485-502.

9. Liang, P., et al., *Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment.* Engineering Applications of Artificial Intelligence, 2022. **115**.

10. Xiao, D., et al., *Fault Diagnosis of Asynchronous Motors Based on LSTM Neural Network*, in *Prognostics and System Health Management Conference (PHM-Chongqing)*. 2018.

11. Morlet, J., et al., *Continuous Wavelet Transforms - Part 1: Review and Extensions*, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 1982. p. 234-238.

12. Wang, D., et al., *Sparsity guided empirical wavelet transform for fault diagnosis of rolling element bearings.* Mechanical Systems and Signal Processing, 2018. **101**: p. 292-308.

13. Lecun, Y., et al., *Gradient-based learning applied to document recognition.* Proceedings of the IEEE, 1998. **86**(11): p. 2278-2324.

14. Phung, V.H. and E.J. Rhee, *A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets.* Applied Science 2019.

15. Szegedy, C., et al., *Going Deeper with Convolutions*, in *Computer Vision and Pattern Recognition (CVPR)*. 2014. p. 1-9.

16. Hochreiter, S. and J. Schmidhuber, *Long Short-Term Memory.* Neural Computation, 1997. **9**(8): p. 1735-1780.