

Deep Learning for Industrial Fault Detection Using Time Frequency Analysis

Ameer Ali Shaalan, Wafa Mefteh and Ali Mohsen Frihida

Robotics, Computing and Complex Systems, National Engineering School of Tunis, University of Tunis-Manar, 1002 Tunis, Tunisia

ameeralishaalan.alkhaykanee@enit.utm.tn, wafa.mefteh@enit.utm.tn, ali.frihida@enit.utm.tn.

Keywords: Fault Diagnosis; CWT; Time-Frequency Analysis; Convolutional Neural Network; ResNet18; Industrial Monitoring.

Abstract: This paper presents an automated multi-class industrial fault diagnostic methodology based on the Continuous Wavelet Transform (CWT) and a fine-tuned ResNet18 deep residual network. Normalised one-dimensional time-series signals for eight fault scenarios were transformed into 64 x 64 time-frequency scalograms using the Morlet wavelet. These single-channel scalograms were used to fine-tune a ResNet18 model modified to accept grayscale input and classify eight categories. The proposed approach was tested on a 1,000-sample dataset divided into three sets training, validation, and test. The experimental results showed an overall test accuracy of 92%, with a weighted precision of 0.93, a recall of 0.92, and an F1-score of 0.92. The model performed well in classification for the Bearing, Blocking, Bearing & Blocking, and No-Fault conditions, but had somewhat worse recall for compound fault classes. These results show the power and reliability of combining CWT-based time-frequency representation with deep residual learning that can be applied in automated industrial condition monitoring.

1 INTRODUCTION

Fault diagnosis is crucial in industrial systems for reducing unplanned delays, enhancing dependability, and ensuring operational safety. Conventional diagnostic approaches use statistical features in the time domain along with frequency-domain analysis methods like Fast Fourier Transform (FFT) and envelope analysis [1], [2]. However, although being computationally efficient and simpler to realize, these approaches are often poor in capturing the non-stationary and transient nature more frequently appeared in the industrial signals of real world [3].

Numerous studies using time-frequency (TF) analysis methods for fault identification have been conducted to overpass these constraints. To receive localized representations of signals' frequency contents over time, using Continuous Wavelet Transform (CWT), Wavelet Packet Transform (WPT) and Short-Time Fourier Transform (STFT) [4].

Due to its multi-resolution characteristic and great time-frequency localization, the continuous wavelet transform (CWT) shows great potential in

characterizing complex and transient fault signatures [5], [6].

With these developments in TF analysis, the merging of the synthesized TF representations with machine learning and deep learning systems has started to receive more interest, automating the feature extraction process and fault classification. Similar previous methods that applied hand-crafted TF features were either combined with random forests and support vector machines (SVMs) [7] to improve diagnostic accuracy but still did not consider the entire integrated levels of interactions for biomarker identification.

The enormous dimensionality of TF data is a common challenge for these approaches, and they still depend significantly on human feature engineering [8]. Advancements in deep learning over the last few years have enabled the development of fault-diagnostic frameworks that can learn discriminative features from TF images.

Because of their effectiveness in representing temporal dependencies in sequential data, recurrent neural networks (RNNs) such as bidirectional LSTM (BiLSTM) and long short-term memory (LSTM) models have seen extensive use [9], [10]. However,

these models may miss spatial features in two-dimensional TF representations, as they focus on temporal interactions [11].

As a result, CNNs have become a formidable substitute, learning hierarchical spatial properties from TF representations of pictures. Convolutional neural networks (CNNs) trained on scalograms, and spectrograms have shown to be useful in many fault classification experiments [12]-[14]. One type of network that uses skip connections to increase training stability and performance is the residual network, also known as ResNets. Despite their benefits, ResNet topologies have not been extensively studied for CWT-based industrial fault diagnosis, particularly for compound and multi-class faults [15].

The research introduces a unified framework for industrial defect diagnostics using deep residual learning with Continuous Wavelet Transform (CWT)-based time-frequency features. Therefore, the one-dimensional industrial time-series signals and non-stationary and transient fault characteristics are illustrated by two-dimensional CWT scalogram. We adapt a pretrained ResNet18 architecture to accept single-channel scalograms and configure for multi-class fault classification. It demonstrates that residual conv nets are capable of learning spatial features which discriminate from time-frequency data.

2 RELATED WORK

Rotating equipment and industrial condition monitoring have been the primary areas of time-frequency-based fault diagnostic research. To obtain non-stationary and transient fault signatures from vibration and acoustic data, CWT-based representations have been widely used in time-frequency analysis [5], [6].

CWT scalograms can accurately capture localized changes of energy due to bearing defects and compound fault conditions because they are characterized by a strong time-frequency localization [12]. Since the CWT obtained representations are limited, investigating their assistance to deep learning models for enhanced diagnostic performance has been the subject of many research efforts. NOTE: Higher classification accuracy have been (achieved by hybrid frameworks that combined CWT and recurrent neural networks (ReN) such as Bidirectional Long Short-Term Memory BiLSTM and Long Short-Term Memory LST [9], [10] by using time frequency information along with temporal dependencies in signal data. However, recurrent-based architectures can face

scaling challenges with high-dimensional time-frequency images and generally a significant number of parameters need tuning [11], which is not applicable in larger-scale industrial system.

In order to tackle these restrictions, direct inference on time-frequency images has been introduced with CNN based methodologies. It is well established that CNNs learn discriminative morphological features when using scalograms and spectrograms, which makes them less sensitive to noise or changes in operating conditions [13], [14]. Additionally, lightweight CNN architectures can provide an excellent trade-off between prediction accuracy and computational cost for real time applications [16].

Recent years have witnessed an increase in interest towards residual CNN architectures to train deeper networks without affecting speed using skip connections. ResNet based approaches for fault diagnostics achieved improved feature representation and generalisation performance [15], [17] Yet hardly any work provided a systematic comparison of the use of pre-trained ResNet architectures in multi-class and compound fault diagnostic context with adaptation to single-channel CWT-based representations [18]. Alam et al. [19] developed a method for classifying multimodal bearing faults by using a 1D convolutional neural network with transfer learning in conjunction with working under different scenarios. A hybrid differential evolution-deep learning framework was presented by Silva et al. [20] to identify defects in induction motors.

3 METHODOLOGIE

Proposed fault classification system time-frequency-based signal enhancement using a CWT and its corresponding ResNet18 variant deep residual learning model is described in this section. The overall procedure of the proposed method consists four central steps: signal acquisition, time-frequency conversion, deep feature extraction and fault recognition.

3.1 Dataset description

This study is based on the data set used originally that has 8 classes: All, Bearing, Blocking, Bearing & Blocking, Leak, Bearing & Leak and Leak & Blocking and No Fault. Each fault class is assigned 125 samples, which achieves an equivalent number of samples in each class for unsupervised training and prevents bias in classification.

While the dataset contains only 1,000 samples, which may be considered small by comparison to large scale image datasets such as ImageNet, it should be noted that industrial fault diagnosis datasets are often limited due to the challenge in obtaining labelled fault data under specific operating conditions. Additionally, a ResNet18 pre-trained model applied via transfer learning resulted in feature extraction with relatively few data set. We employed a stratified data split with balanced class distribution to obtain valid evaluation results. Future work will use a larger dataset to improve generalisation.

3.2 Signal Representation and Preprocessing

In analyzing the one-dimensional discrete-time series signal $x(t) \in \mathbb{R}, 16$ where $t = 1, 2 \dots N$ is the time index and N is the signal length [14] that has been taken from industrial systems under control. We also specify many signals for the system capabilities. Raw signals (390 kHz) are normalized before feature extraction, both to mitigate scale variation and to achieve better numerical stability of training within the network. Eq. 8 is utilized to perform min-max normalization on every signal. 1:

$$x_{\text{norm}}(t) = \frac{x(t) - \min(x)}{\max(x) - \min(x)}. \quad (1)$$

Normalization of input signals is performed before the training process, which reduces bias when converting from amplitudinal characteristics to other properties and improves convergence to optimized values.

3.3 Continuous Wavelet Transform (CWT)

To effectively capture the non-stationary and transient characteristics of industrial time-series signals, CWT is employed for time-frequency analysis. Unlike Fourier-based methods, CWT provides a multi-resolution representation that preserves both temporal and spectral information.

The CWT of a signal $x(t)$ is defined as (2):

$$W(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt. \quad (2)$$

where a denotes the scale parameter controlling frequency resolution, b represents the translation parameter governing time localization, $\psi(\cdot)$ is the mother wavelet, and denotes complex conjugation. Therefore, the Morlet wavelet is chosen as basic

wavelet in this study for its balanced sensitivity to time and frequency localization, making it ideal for the analysis of signal characteristics related to faults. The amplitude of the resultant wavelet coefficients is calculated to form a two-dimensional scalogram, (3):

$$S(a, b) = |W(a, b)|. \quad (3)$$

Scaling is defined as the inverse nature of coincidence - a property whereby different levels of detail are balanced with their physical representation in real-time. Each scalogram contains information about how much signal energy is distributed at various scales and time instants. To align with the convolutional neural network input, all scalograms are resized to a standard resolution of pixels and interpreted as single-channel gray-scale images.

3.4 Deep Residual Network Architecture (ResNet18)

The extracted CWT scalograms are used as inputs to a deep convolutional neural network based on ResNet18 architecture. ResNet18 incorporates residual connections that alleviate the vanishing gradient problem and facilitate stable training of deeper networks. Given an input scalogram image $\mathbf{X} \in \mathbb{R}^{64 \times 64 \times 1}$, The residual learning mechanism can be expressed a (4):

$$y = \mathcal{F}(\mathbf{X}, \mathbf{W}) + \mathbf{X}. \quad (4)$$

where $\mathcal{F}()$ denotes the residual mapping learned by stacked convolutional layers and $\mathbf{W}$ represents the learnable parameters. Since the original ResNet18 architecture is designed for three-channel RGB images, the first convolutional layer is modified to accept single-channel input while preserving the remaining network structure.

Additionally, the final fully connected layer is replaced with a new dense layer corresponding to the number of fault classes.

3.5 Training Strategy and Classification

We fine-tune the modified ResNet18 model based on the CWT scalograms in a supervised learning setting. The network is trained to minimize the categorical cross-entropy loss function denoted as Eq.5:

$$\mathcal{L} = - \sum_{i=1}^C y_i \log(\hat{y}_i). \quad (5)$$

where C is the number of fault classes, y_i is the ground-truth label and $\hat{y}_i$ is the predicted probability of class i .

Table 1: Basic configuration of the improved ResNet18 architecture.

Stage	Layer Type	Output Size	Kernel Size	Stride	Padding	Filters
Input	-	1×64×64	-	-	-	-
Conv1	Convolution	64×32×32	7×7	2	3	64
Pool1	Max Pooling	64×16×16	3×3	2	1	-
Block1	Residual Block ×2	64×16×16	3×3	1	1	64
Block2	Residual Block ×2	128×8×8	3×3	2	1	128
Block3	Residual Block ×2	256×4×4	3×3	2	1	256
Block4	Residual Block ×2	512×2×2	3×3	2	1	512
GAP	Global Avg Pool	512×1×1	-	-	-	-
FC	Fully Connected	8	-	-	-	8

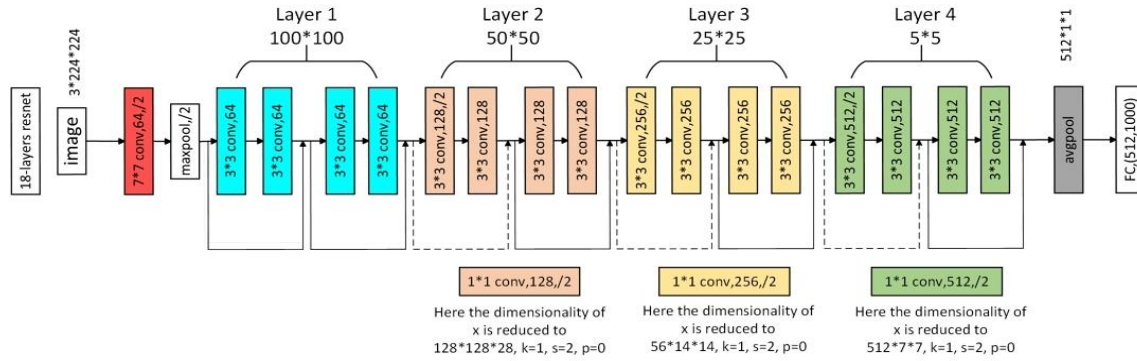


Figure 1: ResNet-18 layers architecture [21].

We use an adaptive optimization algorithm called Adam to update the parameters during training. The dataset is divided into training, validation and test subsets to evaluate the generalization ability of the proposed framework. SoftMax activation function is used at the output layer to obtain the final classification output as (6):

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}}. \quad (6)$$

where z_i denotes the logit value associated with class i .

3.6 Performance Evaluation Metrics

He et al. [22] originally proposed the ResNet18 deep residual network, which serves as the backbone of the proposed architecture. ResNet18 was selected due to its lightweight structure, stable convergence characteristics, and demonstrated effectiveness in image classification applications.

The original ResNet18 architecture was designed for 224×224 RGB images with three input channels. In this study, however, the input data consist of 64×64 single-channel grayscale continuous wavelet transform (CWT) scalograms. Therefore, several modifications were introduced to adapt the

architecture to the characteristics of the input data, as summarized in Table 1.

To comprehensively evaluate the classification performance of the proposed model, a multi-class confusion matrix was generated. The confusion matrix provides detailed information about the relationship between actual and predicted class labels, enabling a class-wise analysis of classification results. Based on the confusion matrix, several widely used performance metrics, including Accuracy, Precision, Recall, and F1-score, were computed [21]. These metrics provide complementary insights into the effectiveness of the model and facilitate a detailed assessment of both overall and class-specific performance.

The evaluation framework combines confusion matrix analysis with the aforementioned classification metrics to provide a comprehensive assessment of the proposed fault diagnosis algorithm. This approach enables the identification of strengths and weaknesses across different fault categories, particularly in cases involving compound faults and classes with similar characteristics. Following the adaptation of the ResNet18 architecture, the next stage involves data organization and preprocessing, as illustrated in Figure 1.

Input: Raw time-series signal $x(t)$.
Output: predicted fault category.

- 1) Acquire raw time-series signal from monitoring sensor.
- 2) Apply min-max normalization to standardize signal amplitude.
- 3) Compute Continuous Wavelet Transform (CWT) using Morlet wavelet.
- 4) Generate magnitude scalogram $S(a, b)$.
- 5) Resize scalogram to 64×64 pixels.
- 6) Feed scalogram into modified ResNet18 network.
- 7) Perform forward propagation.
- 8) Compute categorical cross-entropy loss.
- 9) Update network weights using Adam optimizer.
- 10) Apply SoftMax activation to obtain class probabilities.

Algorithm 1: Proposed automated fault diagnosis framework.

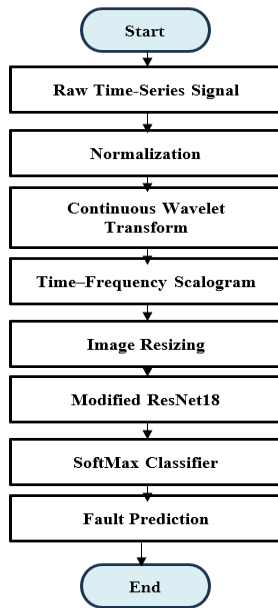


Figure 2: Flowchart for the proposed CWT-ResNet18-based multi-class industrial fault diagnostic framework.

To specify the processing flow of the proposed fault diagnosis system, Algorithm 1 summarizes the whole processing procedure displayed in Figure 2.

4 RESULTS AND DISCUSSIONS

CWT is an essential preprocessing step that transforms one-dimensional time-series data into a two-dimensional format, making it suitable as input to Convolutional Neural ResNet18 architecture was used in this study. In this approach, the raw time-

series signal is utilized; that is, it uses the actual measurements taken at different points in time for each instance of a fault.

The continuous wavelet transform (CWT) convolves the time series with a "mother wavelet" function that has been rescaled and translated. The CWT applies wavelets, which are oscillatory functions localized in space rather than the more commonly used Fourier analysis that represents a signal as a superposition of sinusoidal components through long-time intervals.

Wavelets allow different signal components to be observed at other frequencies. Smaller wavelet scales correspond to high-frequency components and short-lived details of the time series, while larger wavelet scales correspond to low-frequency trends and coarser structures. The flipping of the wavelet along time within the whole time series indicates when each frequency component occurs.

Each CWT gives you a set of values that represent how well the scaled and translated wavelet fits with the time series segment. These figures create a two-dimensional matrix. One dimension is time, another scale (or like frequency). This 2D image, sometimes called a scalogram or heatmap, does a fantastic job of showing how the original signal changes over time and frequency. The resulting scalogram provides a two-dimensional time-frequency representation suitable for convolutional feature extraction. CNN then takes this "image" as input. It learns time-frequency-domain patterns that represent different fault conditions by leveraging its ability to extract spatial features. The CWT essentially converts the time-based aspects of a time series into a format that is easier to visualize and highlights elements unique to both time and frequency.

The scalograms of the average CWT reveals characteristic patterns for each fault scenario (All, Bearing, Bearing & Blocking, Blocking, Leak, Leak & Blocking and No Fault) presented in this study as shown in Figure 2. Scalograms have a horizontal time axis, vertical axis for the wavelet scale (inverse of frequency), and color intensity metric for the magnitude of the wavelet coefficients. Various fault type trends appear in time-frequency domain. For some scales, there is an energetic evicition revealing the transient and periodic part of each signal which corresponds to mechanised rapture like Blocking, Bearing and their associates.

On the contrary, no-fault condition indicates uniformity of energy and less swathes of high

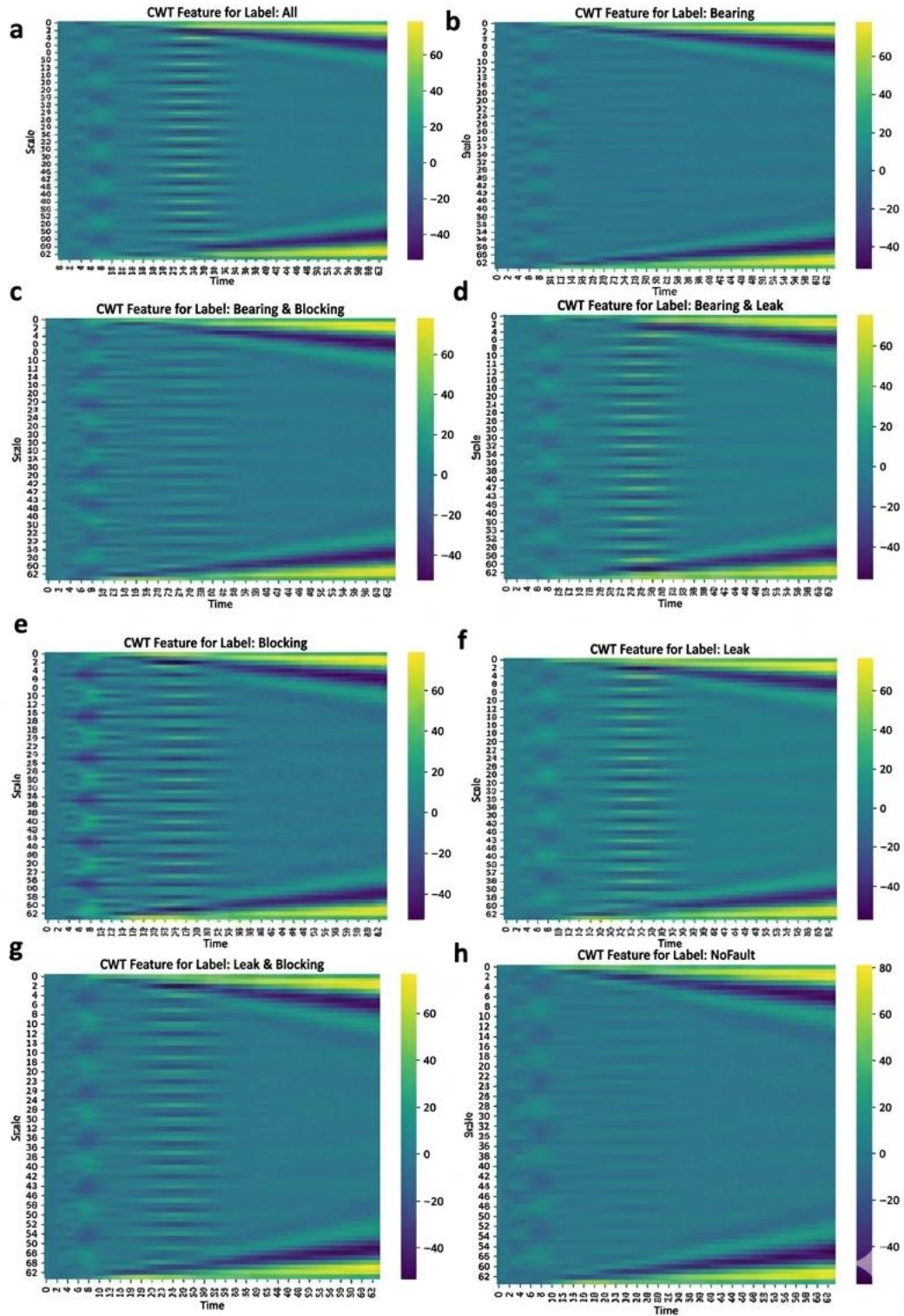


Figure 3: CWT scalograms of different faults conditions are shown, to obtain various time-frequency energy distributions of signal used as input feature for ResNet18 based fault classification model.

intensity shows a stable system. Complex energy patterns distributed on different scales reflect the superposition property of various faults, as demonstrated by compound fault embodiments like Bearing & Leak and Leak & Blocking. A potential reason for the larger misjudgment of some fault classes in condition matrix result is that compound failures own similar time-frequency features. This demonstrates how CWT is capable of automatically classify potential faults by extracting discriminative time-frequency features under different operational conditions and how the ResNet18 can take advantage of these patterns.

Figure 3 represents CWT scalograms for the eight fault scenarios. Each graphic depicts the time-frequency energy distribution of the corresponding signal. Different fault types produce distinct patterns, demonstrating that CWT successfully captures discriminative transient and frequency-dependent properties.

The fine-tuned ResNet18 model yielded a 92% overall classification accuracy on the test dataset, and was determined to be an adequate diagnostic tool. The recall and F1-score averages over the classes are both ~ 0.92 , showing that our model always classifies faults across all classes while accounting for class support. The model has very high precision and recall, with F1-scores close to 1.00 for the Bearing, Bearing & Blocking, Blocking, and No-Fault classes. That suggests the model is very good at identifying these fault conditions, and it doesn't confuse them with other classes too often.

However, the All, Bearing & Leak, Leak and Leak & Blocking classes show a slight performance decrease. In particular, the low recall of the All and Leak & Blocking classes suggests that there are some misclassifications for these fault situations. This tendency is demonstrated by the confusion matrix by having non-zero items off-diagonal in the appropriate rows. For the Bearing & Leak and Leak classes, lower precision indicates that some of these faults are incorrectly forecast. The confusion matrix and the classification report are in agreement. The dominating diagonal elements indicate that for the most part, classes were predicted correctly. The classes confusions are revealed as distinct off-diagonal entries to window/frame errors, by which we can realize the types of faults that have a similar time-frequency features. While the ResNet18 model does well across most fault types, separability between

some compound faults can be improved as shown (i.e., right-hand Fig. 3). The overall classification performance of ResNet18 on the test dataset weighted precision, recall, and F1-score and accuracy as in Table 2. The confusion matrix in Figure 4 demonstrates the effectiveness of the proposed fault classification approach.

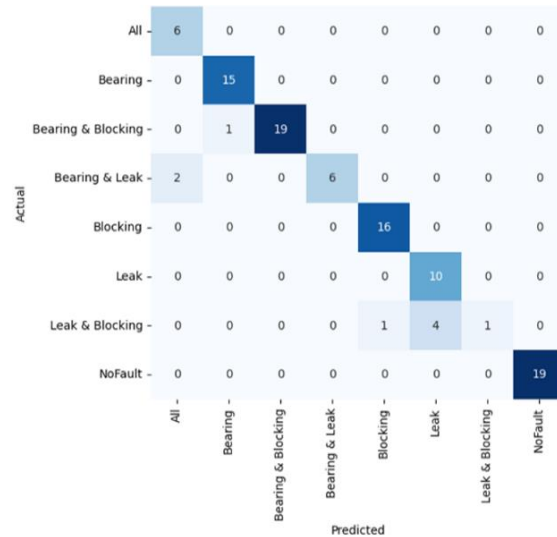


Figure 4: Confusion matrix of ResNet-18.

5 CONCLUSIONS

In this study, it was described a CWT-ResNet18 architecture for automated multi-class industrial fault diagnosis. The suggested approach converts one-dimensional time-series signals into 64×64 time-frequency scalograms and features the use of a modified ResNet18 model for feature extraction and classification. Model, which achieved 92% classification accuracy on test dataset and performed much better in majority of the fault categories. Compound fault classes showed a slight performance decrease due to overlapping time-frequency characteristics. The findings show that combining time-frequency analysis and deep residual learning yields a reliable solution for industrial condition monitoring. Future studies will expand the dataset, explore different wavelet functions, and employ advanced data augmentation and ensemble learning methods to improve compound fault discrimination.

Table 2: Summarizes the overall performance of ResNet18-based fault diagnosis framework.

Precision (weighted avg)	Recall (weighted avg)	F1-Score (weighted avg)	Accuracy	Test data
0.93	0.92	0.92	0.92	100

REFERENCES

- [1] X. Ye, X. Ma, and C. Wen, "Rotating machinery fault diagnosis method by combining time-frequency domain features and CNN knowledge transfer," *Sensors*, vol. 21, no. 24, p. 8168, 2021, doi: 10.3390/s21248168.
- [2] G. Xu, M. Liu, Z. Zhuofu, D. Söffker, and W. Shen, "Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 19, no. 5, p. 1088, 2019, doi: 10.3390/s19051088.
- [3] M. F. Siddique, F. Saleem, M. Umar, C. H. Kim, and J.-M. Kim, "A hybrid deep learning approach for bearing fault diagnosis using continuous wavelet transform and attention-enhanced spatiotemporal feature extraction," *Sensors*, vol. 25, no. 9, p. 2712, 2025, doi: 10.3390/s25092712.
- [4] M. F. Siddique, S. Ullah, and J. M. Kim, "A Deep Learning Approach for Fault Diagnosis in Centrifugal Pumps through Wavelet Coherent Analysis and S-Transform Scalograms with CNN-KAN," *Computers, Materials & Continua*, vol. 84, no. 2, 2025.
- [5] Y. Yan, T. Liu, X. Ye, Q. Jing, and Y. Dai, "Rotating Machinery Fault Diagnosis Based on a Novel Lightweight Convolutional Neural Network," *PLoS ONE*, vol. 16, no. 8, p. e0256287, 2021, doi: 10.1371/journal.pone.0256287.
- [6] Q. Zhang et al., "Intelligent fault diagnosis of rotating machinery using time-frequency images and hybrid CNN architecture," *Scientific Reports*, vol. 15, p. 12344, 2025, doi: 10.1038/s41598-025-95895-x.
- [7] J. Yuan et al., "Image representation of vibration signals and its fault diagnosis using CNN," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2025, doi: 10.1177/09544062241284465.
- [8] Q. Verstraete et al., "Deep learning enabled fault diagnosis using time-frequency representations," *Mathematical Problems in Engineering*, vol. 2017, 2017, doi: 10.1155/2017/5067651.
- [9] X. Zhu et al., "Rotor fault diagnosis using a convolutional neural network," *Mechanical Systems and Signal Processing*, vol. 125, 2019, doi: 10.1016/j.ymssp.2019.01.039.
- [10] L. Pan, X. Zhang, Z. Jiang, et al., "Real-time intelligent diagnosis of co-frequency vibration faults in rotating machinery," *Chinese Journal of Mechanical Engineering*, vol. 37, p. 41, 2024, doi: 10.1186/s10033-024-01021-9.
- [11] H. Wang, H. Wang, and X. Tang, "A review of deep learning in rotating machinery fault diagnosis and its prospects for port applications," *Applied Sciences*, vol. 15, no. 21, p. 11303, 2025, doi: 10.3390/app152111303.
- [12] L. Zhang et al., "An imbalanced fault diagnosis method based on time-frequency feature oversampling and CNN," *International Journal of Machine Learning*, vol. 13, 2022.
- [13] S. Liu et al., "Intelligent fault diagnosis of rotating machine via hybrid deep learning," *Knowledge-Based Systems*, 2025.
- [14] X. Li, W. Zhang, Q. Ding, and J. Q. Sun, "Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation," *Journal of Intelligent Manufacturing*, vol. 31, no. 2, pp. 433-452, 2020.
- [15] S. Tang, S. Yuan, and Y. Zhu, "Deep learning-based intelligent fault diagnosis methods toward rotating machinery," *IEEE Access*, vol. 8, pp. 9335-9346, 2019.
- [16] T. Luo, M. Qiu, Z. Wu, Z. Zhao, and D. Zhang, "Bearing Fault Diagnosis Based on Multi-scale Spectral Images and Convolutional Neural Network," *Journal of Vibroengineering*, vol. 27, no. 6, pp. 1012-1027, 2025, doi: 10.21595/jve.2025.24934.
- [17] H. Lee and J. Yu, "A fault detection framework for rotating machinery with a spectrogram and convolutional autoencoder," *Applied Sciences*, vol. 15, no. 14, p. 7698, 2025.
- [18] J. Zhanget al., "Fault diagnosis for rotating machinery based on deep learning integrating multi-domain features," *Measurement Science and Technology*, 2024, doi: 10.1088/1361-6501/adc774d.
- [19] T. E. Alam, M. M. Ahsan, and S. Raman, "Multimodal bearing fault classification under variable conditions: A 1D CNN with transfer learning," *Machine Learning with Applications*, vol. 21, p. 100682, 2025.
- [20] R. M. Silva, A. L. de Oliveira Vitor, M. F. Castoldi, A. Goedtel, and W. A. Souza, "Optimized Time-Frequency Analysis for Induction Motor Fault Detection Using Hybrid Differential Evolution and Deep Learning Techniques," *International Journal of Adaptive Control and Signal Processing*, 2025, doi: 10.1002/acs.4075.
- [21] C. Wang, "Federated Learning with ResNet-18 for Medical Image Diagnosis," in *Proceedings of the 2023 8th International Conference on Multimedia Systems and Signal Processing (ICMSSP 2023)*, Shenzhen, China, pp. 1-6, ACM, New York, NY, USA, 2023, [Online]. Available: <https://doi.org/10.1145/3613917.3613922>.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.