A Novel Approach for Fault Diagnosis in Mechanical Systems Using Time-Frequency Analysis and Unsupervised Learning

Kuo-Hao Li[1] and Yao-Chi Tang[2,*]

[1]Department of Engineering Science and Ocean Engineering, National Taiwan University, Taipei City 106319, Taiwan
[2]Department of Systems Engineering and Naval Architecture, National Taiwan Ocean University, Keelung City 202301, Taiwan
Email: d10525008@ntu.edu.tw (K.-H.L.); tom@mail.ntou.edu.tw (Y-C.T.)
[*]Corresponding author

Manuscript received April 10, 2023; revised May 5, 2023; accepted August 23, 2023; published February 6, 2024

***Abstract*—The development of a new method for fault diagnosis in mechanical systems is a critical field of research due to the increasing demand for machine reliability and maintenance efficiency. In this study, a novel approach to fault diagnosis using time-frequency analysis and unsupervised learning techniques is proposed. Firstly, the proposed method converts the vibration signal into a time-frequency domain signal using the Short-Time Fourier Transform (STFT) and integrates it with respect to time to obtain the Marginal Frequency (MF). The Area Under the Frequency Curve (AUFC) is calculated and an unsupervised 1D K-means clustering algorithm is used to cluster the feature vectors. Each cluster is assigned a normal or failure state and the maximum value of the normal region is used as the threshold for failure detection. The method is tested on a set of vibration data from normal and failed bearings, and the results demonstrate the effectiveness and robustness of the proposed approach for fault detection in different bearings. The proposed approach represents a promising solution for fault diagnosis in mechanical systems, which can significantly improve the reliability and maintainability of machinery. The combination of time-frequency analysis and unsupervised learning techniques provides a powerful tool for fault detection in mechanical systems.**

***Keywords*—fault diagnosis, K-means clustering algorithm, marginal frequency, time-frequency analysis**

I. INTRODUCTION

Fault diagnosis in mechanical systems is an important field in the research, driven by the increasing demand for reliable machinery and efficient maintenance. The ability for the quickly and accurately detect faults in mechanical systems is essential to prevent catastrophic failures and minimise downtime. Therefore, there is a requirement for effective and efficient fault diagnosis methods. Traditional methods of fault diagnosis in mechanical systems rely on manual inspection and judgement based on experience. However, these methods waste time, costly and prone to error. With the rapid advancement of sensor and signal processing technologies, machine learning-based fault diagnosis methods have increased their high accuracy, efficiency and automation. Various methods have been proposed for fault diagnosis in mechanical systems, such as Root-Mean-Square (RMS) [1–3], Kurtosis (KU) [4–6] and envelope analysis [7]. However, these methods have limitations in handling complex signals and detecting incipient faults. To overcome these limitations, this study proposes a novel approach to fault diagnosis using time-frequency analysis and unsupervised learning techniques. Firstly, the proposed method converts the time-domain vibration signals to a time-frequency domain signal using the STFT [8], which provides a more complete representation of the signal. Then, by integrating the time-frequency signal with respect to time, the MF is got and the AUFC is calculated as a feature vector. An unsupervised 1D K-means clustering algorithm [9] is used to cluster the feature vectors and each cluster is assigned a normal or failure state. The maximum value of the normal region is regarded the threshold for failure detection. The proposed approach offers a unique contribution to fault diagnosis in mechanical systems by combining time-frequency analysis and unsupervised learning techniques. This innovative integration enhances the signal representation, facilitates effective feature extraction and classification, and demonstrates promising potential for real-world applications. With further development and optimization, the proposed method has the potential to become a practical and reliable solution for fault diagnosis, significantly improving the reliability and maintainability of machinery. In conclusion, the proposed approach presents a novel methodology that addresses the limitations of previous research. By leveraging the strengths of time-frequency analysis and unsupervised learning, it offers an innovative and powerful tool for fault detection in mechanical systems. The study contributes to the development of advanced fault diagnosis techniques and highlights the potential of time-frequency analysis and unsupervised learning in mechanical system applications.

II. METHODOLOGY

This section introduces the relevant theories used in the diagnostic analysis process. The main method in this study is Short-Time Fourier Transform (STFT), which provides time-frequency analysis results. The Marginal Frequency (MF) is obtained by integrating the STFT results for time domain. The Area Under the Frequency Curve (AUFC) is used as the diagnostic feature, and a non-supervised learning algorithm, K-Means, is used to automatically classify the signal into normal or faulty categories. This approach allows for the determination of whether the tested signal has experienced a fault.

A. STFT

The use of the STFT has become increasingly popular in recent years for the analysis of time domain signals. Unlike the Fourier transform, which can only determine the energy-frequency variation, the STFT takes into account the

DOI: 10.7763/IJMO.2024.V14.843

time variables in the energy-frequency magnitude, allowing analysis of the frequency variation of non-linear or non-steady-state vibration signals. The STFT principle involves multiplying the input signal by a window function and dividing long time signals into many shorter signals with the same time span. This allows the Fourier transform to be performed on specific local short-term signals as the window function moves with the time axis, producing a complex function representing the magnitude of the signals as time and frequency change. The main equation of the STFT [10], shown in Eq. (1), involves the integration of a window function and a signal, and is an essential tool for bearing fault analysis.

$$STFT\left[x(t)\right] \equiv X(\tau,\omega) = \int_{-\infty}^{\infty} x(t)\omega(t-\tau)e^{-i\omega\tau}dt \qquad (1)$$

The equation presented earlier in the section shows the principle of STFT and its main equation. In this equation, $x(t)$ represents the input signal, $\omega(\tau)$ is a window function, and $X(\tau,\omega)$ is a complex function that represents the signal's magnitude with respect to time and frequency changes. There are different types of window functions available, such as Hamming and Gaussian windows, which are commonly used in STFT. In this study, Hamming windows were used and its equation is given by Eq. (2):

$$w(n) = 0.5\left[1-\cos\left(\frac{2\pi n}{N}\right)\right] = \sin^2\left(\frac{\pi n}{N}\right), \qquad 0 \le n \le N \qquad (2)$$

where $w(n)$ is the value of the Hamming window at index n, N is the window length. The Hamming window is suitable for analyzing non-stationary signals with low side lobes and good frequency resolution.

B. MF

In the analysis of non-stationary bearing vibration signals, the concept of MF is introduced, which is obtained by integrating the time-frequency two-dimensional array along the time axis. The resulting one-dimensional array represents the energy aggregation corresponding to each frequency and can be used to identify the optimum fault characteristic frequency. The MF provides a more accurate representation of the frequency characteristics of the signal because it takes into account the time variation of the frequency energy. In this study, the MF was obtained by STFT analysis of the time-domain signal and used as a diagnostic feature for fault detection. The mathematical expression for the MF calculation is shown in Eq. (3).

$$MF(\omega) = \int_{-\infty}^{\infty} X(t,\omega)dt \qquad (3)$$

C. K-means Cluster Analysis

K-means is a widely used unsupervised clustering algorithm that divides input data into clusters with similar characteristics based on their distances from each other. The algorithm is often used in pattern recognition and machine learning applications. The K-means clustering algorithm is a simple and efficient approach that can partition data into distinct groups. It consists of the following steps [11]:

- Initialisation: K signals in a given signal set are randomly selected as the initial cluster centre.
- Allocation: The Euclidean distance $d_{ij} = \left\| x_i - \mu_j \right\|^2$ between each signal and each cluster centre is calculated and each signal is assigned to the nearest cluster.
- Recalculation: The centre of k clusters is recalculated using the following method: where is the number of clusters.
- Iteration: The total mean square error between all signals and the corresponding cluster centre is calculated using . If convergence or the maximum number of iterations is reached, proceed to step 5, otherwise return to step 2.
- Output: Export the cluster result.

The K-means algorithm can effectively classify data and is widely used in various fields. In fault diagnosis, K-means can automatically classify normal and fault signals and determine the optimum threshold for fault detection.

III. Experimental Setup

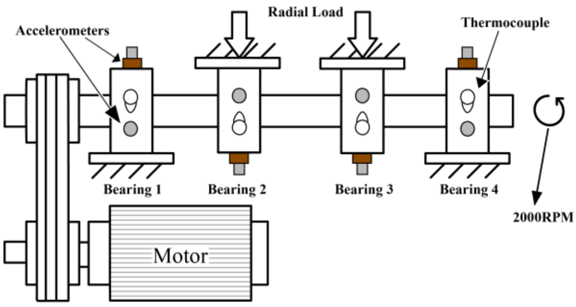

Fig. 1. The bearing test and accelerometer arrangement plan.

To validate the proposed method, the experimental data provided by the NASA Bearing Data Center was adopted as a benchmark dataset for fault diagnosis in mechanical systems. Specifically, the 2nd test data set was used, which includes vibration signals recorded from four bearings installed on a shaft of a test platform, each receiving a radial load of 6,000 lb. The test platform, equipped with an AC motor driving the shaft at a constant speed of 2000 rpm, is shown in Fig. 1. To obtain the vibration signals, an accelerometer is installed on each bearing seat and the signals are collected using a data acquisition system. In addition, the bearings were lubricated during the experiment to maintain their operating conditions. To ensure the safety of the test platform, a magnetic plug is installed in the oil return line to collect metal debris resulting from bearing wear. This allows the operation to be stopped when a certain amount of debris accumulates. Further details of the test platform can be found in the dataset documentation [12, 13]. A set of experimental data consisting of 984 files was selected for this study. Each file represents the data extracted from a measurement taken on the equipment every ten minutes. A one-second vibration signal was recorded each time, with the sampling frequency set to 20 kHz. Each file contains four channels, corresponding to the measured data from the four accelerometers. Each row contains 20,480 points representing the vibration signal sampled at high frequency. The experiment continued until the outer ring of bearing 1 failed, which is a common failure mode in mechanical systems. The failure was detected and verified by visual inspection of the bearing after the experiment. The use of this dataset ensures the reliability and relevance of the results as they are validated against a widely accepted benchmark dataset.

IV. Results and Discussion

A. AUFC

The use of time domain vibration signals alone is not sufficient to accurately diagnose bearing damage. To overcome this problem, the STFT was used to convert the time domain signal to a time-frequency domain signal, which was then integrated over time to obtain the MF for analysis. Specifically, a one second segment was selected for analysis in each ten minute measurement interval. The results of the MF analysis for normal versus normal operation and normal versus faulty operation are shown in Figs. 2 and 3. The horizontal axis in both figures. represents the frequency, while the vertical axis represents the corresponding energy magnitude. Fig. 2 shows a comparison of two vibration signals from normal bearings, where a significant overlap in energy magnitudes is observed. In contrast, Fig. 3 shows a comparison between a normal vibration signal and a signal from a faulty bearing in the later stage of the experiment, where a clear difference in energy magnitudes is observed between normal and faulty signals. It should be noted that there may be some variation in energy magnitudes at lower frequencies due to differences in bearing placement and contact friction. In addition, it is common for signals from damaged bearings to occur at higher frequencies, and in Fig. 3 no significant resonance frequency is observed above 6000 Hz, suggesting that only the AUFC between 2000-6000 Hz should be considered in further analysis.

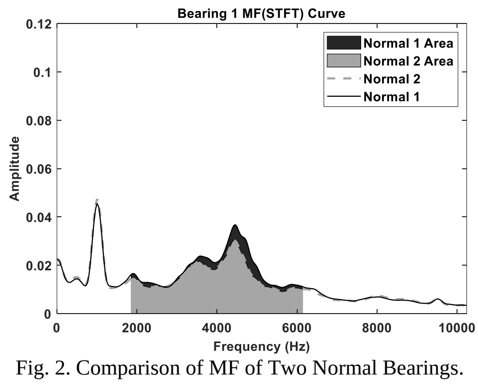

Fig. 2. Comparison of MF of Two Normal Bearings.

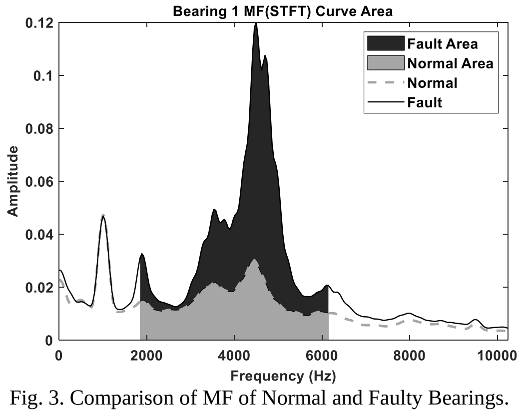

Fig. 3. Comparison of MF of Normal and Faulty Bearings.

In this study, the AUFC was calculated using the trapezoidal integration method. Specifically, the trapezoidal rule was applied to integrate the frequency curve shown in the coordinate system of Fig. 4, which involves calculating the area of the trapezoid ABCD. The calculation is expressed in (4), where a and e are the upper and lower limits of the frequency, respectively. E_i represents the ith trapezoid of the discretised frequency curve, where y_{i-1} and y_i are the upper and lower bases of the trapezoid, respectively, and $\triangle x_i$ is the height of the trapezoid.

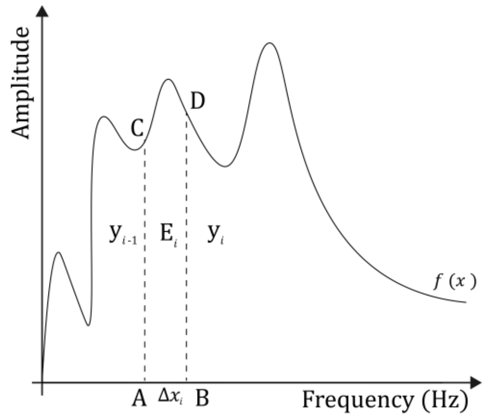

Fig. 4. AUFC calculation.

$$AUFC = \int_a^e f(x)dx = \sum_{i=1}^{n} E_i = \sum_{i=1}^{n} \frac{y_{i-1} + y_i}{2} \triangle x_i \tag{4}$$

B. K-means Resluts

In this study, an unsupervised K-means clustering algorithm was used to automatically classify the vibration signals into two groups, normal and faulty. The proposed method of fault diagnosis is used time-frequency analysis and integrates the resulting MF to obtain the AUFC between 2000-6000 Hz. The AUFC was then used as a feature vector to cluster the data by the K-means algorithm. The K-means clustering algorithm is a popular unsupervised machine learning technique used for data classification. In this study, the 1D K-means algorithm was applied to obtain from the MF integration process. This algorithm clusters the data divided into two groups based on the similarity of their feature vectors. The clustering process is performed iteratively until the cluster centroids don't change significantly. After clustering the data, the next step is to determine the fault detection threshold. This threshold is used to distinguish between normal and faulty states. In this study, the maximum value of the normal region was used as the fault detection threshold.

Fig. 5 shows the classification results by the unsupervised K-means algorithm with 1D clustering. The horizontal axis represents the AUFC values, while the vertical axis represents the classified states. A value of -1 indicates the normal state and a value of 0 indicates the faulty state. The threshold of fault diagnosis is 0.12718, as shown in the figure. If the feature vector of a vibration signal exceeds this threshold, it is classified as a failed state. From the plot, it is clear that the proposed method can effectively distinguish between the normal and faulty states. The majority of the data points fall into either the normal or faulty category, indicating that the proposed method is reliable in fault diagnosis.

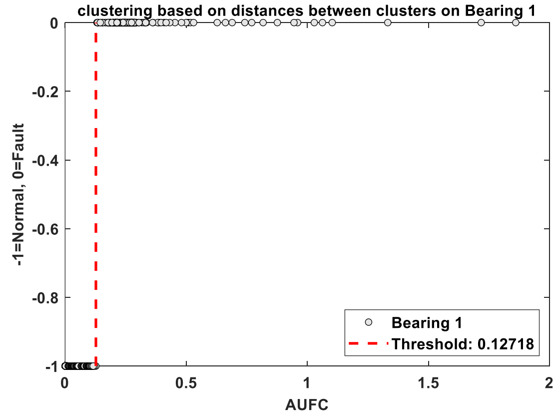

Fig. 5. Classification Results using K-means clustering.

The threshold was obtained by the K-means unsupervised algorithm and 1D clustering was used to automatically classify the vibration signals. The maximum value of the signal features classified as normal was thought the fault diagnosis threshold, which was found to be 0.12718. The effectiveness of this threshold was further analysed by applying it to other bearings in the same test and the results are shown in Fig. 6. The horizontal axis in the figure represents different bearings, each containing 984 measured data points, while the vertical axis represents the AUFC. It can be seen from the figure that at the end of the experiment the threshold only produced false positives in three other bearings. This is because bearing 1 had already suffered severe damage and the other bearings on the same shaft would generate more severe signals as a result of the chain reaction.

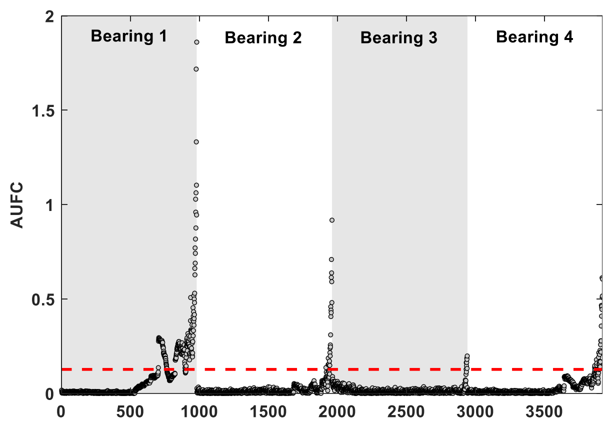

Fig. 6. AUFC analysis of other bearings using the fault diagnosis threshold value.

These results demonstrate that the fault threshold obtained through the K-means algorithm is effective in diagnosing the condition of the bearings. The approach is not only efficient but also highly automated, reducing the workload of the maintenance staff. Moreover, the use of the AUFC allows for more accurate fault diagnosis, a valuable tool for condition monitoring and fault diagnosis of bearings in industrial settings.

V. Conclusion

This study has proposed a novel approach to fault diagnosis in mechanical systems using time-frequency analysis and unsupervised learning techniques. The proposed method was shown to be effective for detecting faults in bearings and has the potential to significantly improve the reliability and maintainability of machinery. The approach involved integrating the STFT and 1D K-means clustering algorithms to cluster feature vectors and assign them a normal or fault state. Using the maximum value of the normal regarded the threshold for fault detection further improved the accuracy of the method. Compared to traditional methods, the proposed approach has several advantages. Firstly, it can detect uncertain faults without prior knowledge of their characteristics. Secondly, it can be used in real operation, which is essential for ensuring the safety and reliability of mechanical systems. Finally, the use of unsupervised learning techniques reduces the need for manual labelling of training data, saving time and effort in the fault diagnosis process. However, the proposed method has limitations that require future research. Firstly, an amount of training data is required to achieve high accuracy, which may prove challenging for certain applications. Secondly, the method may not be suitable for detecting very early-stage faults where the signal is not strong enough to produce a clear MF. Finally, further optimization is required to improve the parameters of the method to achieve even better performance. In conclusion, the proposed approach is a promising solution for fault diagnosis in mechanical systems and has the potential to become a practical and reliable tool for real-world applications. By further refining and optimizing the method, it could contribute to the development of more efficient and reliable machines in the future.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

K. L. conceived and designed the research, collected and analyzed the data, and wrote the manuscript. Y. T. provided critical feedback and helped to revise the manuscript. All authors read and approved the final manuscript.

Acknowledgment

The authors declare that they have no acknowledgments.

References

[1] M. Altaf, T. Akram, M. A. Khan, M. Iqbal, M. M. I. Ch, and C. H. Hsu, "A new statistical features-based approach for bearing fault diagnosis using vibration signals," *Sensors*, vol. 22, no. 5, p. 2012, Mar. 2022.

[2] K. Zhang, Y. Lu, H. Shi, X. Bai, and Z. Yuan, "The fault position identification method for the outer ring of deep groove ball bearing in a wide range," *Measurement*, vol. 203, 111939, Nov. 2022.

[3] C. Xi, J. Yang, D. Zhen, X. Liao, W. Hu, and F. Gu, "Application of combined normalized least mean square and ensemble empirical mode decomposition denoising method in fault diagnosis of rolling bearings," in *Proc. IncoME-VI and TEPEN 2021*, pp. 629–641, Sep. 2022.

[4] B. Chen *et al*., "Integrated early fault diagnosis method based on direct fast iterative filtering decomposition and effective weighted sparseness kurtosis to rolling bearings," *Mechanical Systems and Signal Processing*, vol. 171, 108897, May 2022.

[5] M. Tang, Y. Liao, R. Duan, J. Xue, and X. Zhang, "Bearing fault diagnosis based on the maximum squared-enveloped multipoint kurtosis morphological deconvolution," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[6] L. Hua, X. Wu, T. Liu, and S. Li, "The methodology of modified frequency band envelope kurtosis for bearing fault diagnosis," *IEEE*

Transactions on Industrial Informatics, vol. 19, no. 3, pp. 2856–2865, Mar. 2023.
[7] B. Chen *et al*., "Product envelope spectrum optimization-gram: An enhanced envelope analysis for rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 193, 110270, Jun. 2023.
[8] W. Guoxin, W. Ge, L. Xiuli, and D. Ruilong, "Bearing fault diagnosis method based on STFT image and AlexNet network," *Mechanisms and Machine Science*, pp. 1056–1068, 2023.
[9] X. Zhang, L. He, X. Wang, J. Wang, and P. Cheng, "Transfer fault diagnosis based on local maximum mean difference and K-means," *Computers and Industrial Engineering*, vol. 172, 108568, Oct. 2022. [1]
[10] Q. Zhang and L. Deng, "An intelligent fault diagnosis method of rolling bearings based on short-time fourier transform and convolutional neural network," *Journal of Failure Analysis and Prevention*, Feb. 2023.
[11] X. Cai, L. Cheng, and Q. Yao, "Rolling bearing health state assessment based on K-means and ensemble HMM," in *Proc. 2022 3rd Information Communication Technologies Conference (ICTC)*, May 2022.
[12] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of Sound and Vibration*, vol. 289, no. 4–5, pp. 1066–1090, Feb. 2006.
[13] J. Lee, H. Qiu, G. Yu and J. Lin, "Rexnord technical services," IMS University of Cincinnati. "Bearing Data Set"," NASA Ames Prognostics Data Repository, 2007.