Iranian Society of Acoustics and Vibration
The 14th International Conference on Acoustics and Vibration ISAV2024
11-12 Dec 2024 Karaj - Iran
Kharazmi University

Multi-Scale Attention Entropy for Robust Fault Diagnosis and Fault Severity Estimation in Rotating Machines Without Prior Knowledge

Emadaldin Sh Khoram-Nejad [a], Abdolreza Ohadi [b*], Farshad Almas-Ganj [c]

[a] *PhD Candidate, Acoustics Research Laboratory, Mechanical Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran.*

[b] *Professor, Acoustics Research Laboratory, Mechanical Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran.*

[c] *Associate Professor, Biomedical Engineering Department, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran.*

** Corresponding author e-mail: a_r_ohadi@aut.ac.ir*

Abstract

Assessing the severity levels of faults in rotating machines is a critical endeavour within the industry, owing to the challenging nature of the noisy working environment and the subtle fault characteristics present in the acquired signals. In this study, a new feature extraction method named multi-scale attention entropy (MSAE), which is a combination of the attention entropy (AttnEn) and the multi-scale entropy (MSE) to extract more discriminative features from the signals, is introduced and investigated. A comparison between the MSAE and randomly selected feature vectors built from a set of 32 statistically and probabilistically features, with the same length, is made to show the performance and ability of the MSAE method. The comparison also includes consideration of the feature vector extracted from the multi-scale sample entropy (MSSE), which is the earliest version of the MSE. Subsequently, all ten feature vectors are input into a support vector machine (SVM) classifier for fault diagnosis and estimation of fault severities. Finally, the performance of the methods is compared for two scenarios, fault diagnosis (FD) and fault diagnosis and severity estimation (FD&SE), on two challengeable datasets. The first dataset, the Case Western Reserve University bearing (CWRU) dataset, is a bearing fault dataset, while the second one, the Korea Advanced Institute of Science and Technology (KAIST) dataset, is a rotor-bearing fault dataset. After twenty iterations, the MSAE-SVM model achieved an average FD accuracy of "99.58%±0.57%" for the CWRU dataset and "93.05%±0.66%" for the KAIST dataset. In addition, the FD&SE accuracy of the MSAE-SVM model for CWRU and KAIST datasets were 98.64% $\pm$ 0.68% and 95.75% $\pm$ 0.71%, respectively. According to the accuracy tolerance of the feature vectors results from the MSAE-

SVM, which is lower than those of other feature vectors, the presented model is more robust in testing accuracy. The presented model is also free of prior knowledge classification and presents much higher mean accuracy among other models used for comparison.

Keywords: Fault diagnosis; Severity estimation; Multi-Scale entropy; Attention entropy.

1. Introduction

In the contemporary era, rotating machines are prevalent in various industries, ranging from expansive power plants to modest workshops. Considering safety, it is important to note that no rotating machine in the industry can be deemed fully healthy because of potential faults arising after the manufacturing or assembly. Also, considering its environment and working conditions, some faults may appear in the rotating machine after working for a while. It is arduous and also necessary to accurately diagnose the weak characteristics of faults during the early stages because it reduces maintenance costs and prevents secondary faults from appearing. Hence, the industry places great importance on fault diagnosis and accurately estimating the severity of the faults.

The fault diagnosis and severity estimation procedure primarily depend on the extracted and selected features. The features must be discriminative, which separates the faulty signals with various severities from each other. Traditionally, researchers used time-domain features such as root mean square (RMS), skewness, kurtosis, mean value, and standard deviation (STD) for feature extraction and selection in fault diagnosis of rotating machines [1]–[3]. However, the entropy-based features, which demonstrate the complexity and irregularities of vibrational signals successfully, have been studied in recent years. Entropy was first introduced in 1948 by Shannon and called Shannon entropy [4]. By examining the probability distribution of its states, Shannon entropy provides an estimate of the complexity of a system. After that, developed versions of Shannon entropy and other entropies are presented to reach more discriminative features of non-linear and non-stationary signals. One of the challenging problems of the various types of entropies is that they are almost parameter-dependent and time-consuming for long time series. To deal with these problems, Yang et al. [5] proposed a novel entropy in 2020, referred to as attention entropy (AttnEn). Traditional entropy measures focus on the frequency distribution of all observations in a time series, whereas the AttnEn analyzes the frequency distribution of intervals between key observations. This approach provides faster computation and improved performance. In addition, the AttnEn remains robust regardless of the time series length. In addition, all the entropies indicate the irregularity and complexity of signals from a single time scale, and it limits their further application. To have a feature vector from a signal, one can use multi-scale entropies (MSE). Costa et al. [6] employed MSE to thoroughly analyze the temporal signal complexity across various scales, effectively addressing the challenge of a single entropy in fully capturing the characteristics of faults. The MSE and entropies can be combined to have a better feature vector and more accurate fault diagnosis [7], [8].

In this study, the application and effectiveness of a feature extractor, combining the MSE and the AttnEn, called multi-scale attention entropy (MSAE) in fault diagnosis and severity estimation of rotating machines is investigated. The raw time series are divided into 2048-point signals and the MSAE of them is calculated. Therefore, according to the optimal scale numbers of the MSAE, which have been obtained considering both the time-consuming and the effectiveness of the MSAE on the two datasets, a feature vector is achieved. Eight feature vectors, including some features from a set of 32 commonly used statistical and probabilistically features, are randomly selected for comparison. The number of features selected from a set of 32 features, must be the same as the length of the feature vector results from the MSAE to have a fair comparison. The reason for using such feature vectors is to show that the MSAE is a feature extractor without the need for prior knowledge. In addition, a

comparison is carried out between the MSAE and multi-scale sample entropy (MSSE), as the first type of MSE to show the superiorities of the MSAE feature extractor. Finally, the performance of the MSAE feature vector is compared using a support vector machine (SVM) classifier with other feature vectors for two scenarios, fault diagnosis (FD) and fault diagnosis and severity estimation (FD&SE), on two challenging datasets. The first dataset is the Case Western Reserve University bearing (CWRU) dataset, which is a bearing fault dataset. The second dataset is the Korea Advanced Institute of Science and Technology (KAIST) dataset, which is a rotor-bearing fault dataset.

The paper is constructed as in the second section, the AttnEn, MSE, and then MSAE calculation procedures are fundamentally described. In section 3, the implementation results of the MSAE and other nine feature vectors on two datasets and under two scenarios are presented and compared. Finally, in the fourth section, the results obtained in the previous section are discussed and concluded and the most important activeness of this study are highlighted.

2. Methodology

2.1 Attention Entropy (AttnEn)

The attention entropy has some fundamental differences from other entropies, which make it a parameter-free and effective method [5]. Considering $X = \{x_1, x_2, x_3, \dots, x_n\} = x_i$, where $1 < i < N$, as the original time-series are supposed to calculate the attention entropy of it. It has some extreme points (local minima and local maxima) called key points. This kind of entropy uses the entropy of the probability of time-series as Shannon entropy but pays attention to the key points. One can define a local maxima series (P) with a length of m and a local minima series (Q) with a length of n, as:

$$P = \{p_1, p_2, \dots, p_m\} \tag{1a}$$

$$Q = \{q_1, q_2, \dots, q_r\} \tag{1b}$$

where subscripts m and r are the number of local maxima and local minima points, respectively.

Then, four types of sub-series (I) can be defined as intervals between local maxima to local maxima points I_{max}^{max}, local maxima to local minima points I_{max}^{min}, local minima to local maxima points I_{min}^{max}, and local minima to local minima points I_{min}^{min}, respectively. The sub-series can be defined using the extreme points defined in Eq. (1a) and Eq. (1b) as follows:

$$I_{max}^{max} = \{x_{p_1-p_2}, x_{p_2-p_3}, \dots, x_{p_{m-1}-p_m}\} \tag{2a}$$

$$I_{max}^{min} = \{x_{p_1-q_1}, x_{p_2-q_2}, \dots, x_{p_m-q_r}\} \tag{2b}$$

$$I_{max}^{min} = \{x_{q_1-p_2}, x_{q_2-p_3}, \dots, x_{q_r-p_m}\} \tag{2c}$$

$$I_{min}^{min} = \{x_{q_1-q_2}, x_{q_2-q_3}, \dots, x_{q_{r-1}-q_r}\} \tag{2d}$$

It is assumed that $p_1 > q_1$. However, it is possible that $x_{p_1-q_1}$, $x_{q_1-p_1}$, $x_{p_m-q_r}$, or $x_{q_r-p_m}$ may be empty for different time series if the first peak in amplitude is local minima, local maxima, $q_1 < p_1$, or $p_1 < q_1$, respectively.

Using these four sub-series, the probability of each element appearing in the set based on the number of appearances is calculated named $f = \{f_{max}^{max}, f_{max}^{min}, f_{min}^{max}, f_{min}^{min}\}$. Next, the Shannon entropy, $H(f)$, is calculated using Eq. (3) for each interval.

$$H(f) = -\sum_{\mathrm{i}} f_i \log_2(f_i) \tag{3}$$

Finally, the average of four calculated entropies is called attention entropy. An example of the AttnEn implementation is illustrated in Figure 1.

$$AttnEn = \frac{H_{max}^{max} + H_{max}^{min} + H_{min}^{max} + H_{min}^{min}}{4} \tag{4}$$

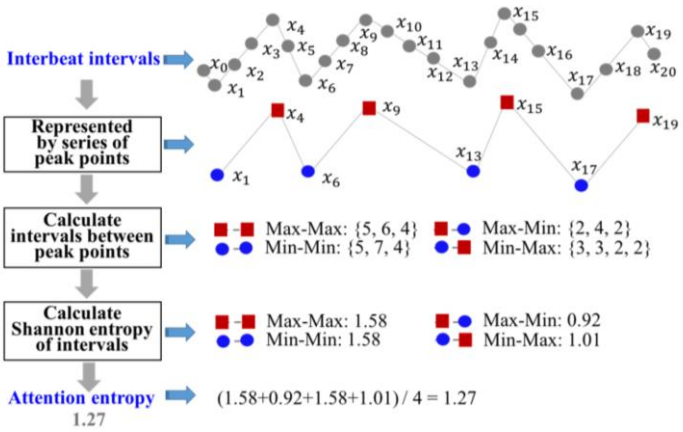


Figure 1. An example of the calculation procedure of the AttnEn for a time series [5].

2.2 Multi-scale attention entropy (MSAE)

The multi-scale entropy (MSE) analysis is a technique to evaluate the complexity and regularity of a signal at multiple time scales [6]. It provides insights into the dynamic temporal fluctuations of the information encoded in a signal and offers additional useful information that the conventional single-value entropy measure fails to capture. The MSE analysis iterates over two steps for each specified scale factor τ: signal graining and entropy calculation. The signal graining procedure for each scale defines different strategies to consider sub-time series with different resolutions. It is noteworthy that by increasing the scale factor, the resolution of the signal is decreased.

The first step of the conventional MSE analysis is to segment the signal into non-overlapping-grained sequences for different temporal scales. A traditional graining method for MSE is the coarse-graining method. Given a signal $x_i = \{x_1, x_2, x_3, \dots, x_n\}$, $for\ i = 1,2, \dots, N$, the coarse-grained signal for scale factor τ ($\tau \in N$), denoted as $x_{g,j}^{\tau}$ and can be calculated by averaging all the data points within the j-th graining window, as shown in Eq. (5). The subscript cg, shows that the signal is coarse-grained, the subscript j is a counter for data point in the grained signal, and the superscript τ shows the grained signal's time scale.

$$x_{cg,j}^{\tau} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i\,, \qquad 1 \le j \le \frac{N}{\tau} \tag{5}$$

In coarse-graining, the averaging window method size is fixed and does not account for variations in local data characteristics. This can be a limitation if the data has non-uniform distributions or trends. One can use the modified-graining method instead of the coarse-graining method as follows:

$$x_{mg,j}^{\tau} = \frac{1}{\tau} \sum_{k=0}^{\tau-1} x_{j-k}\,, \qquad 1 \le j \le N - \tau + 1 \tag{6}$$

where subscript mg refers to the modified-graining method.

It can be obvious from comparing Eq. (5) and Eq. (6) that the length of the grained signal obtained from the modified-graining method is longer compared to the coarse-graining method, but still reduces compared to the original. The length of the signal in the coarse- and the modified-graining methods by increasing time scales from $\tau = 1, 2, 3, 4, \dots$ is $N, \frac{N}{2}, \frac{N}{3}, \frac{N}{4}, \dots$ and $N, N-1, N-2, N-3, \dots$, respectively. Therefore, the modified-graining method offers a more gradual reduction in signal

length and potentially better handling of edge effects due to the sliding window approach. This can be beneficial for capturing local variations and consequently, can be more compatible with attention entropy.

Another difference between the two graining methods is the contributions of each data point of the main signal in the grained sub-signals. Consider $x_i = \{x_1, x_2, x_3, \dots, x_n\},\ for\ i = 1,2, \dots, N$, implementing the coarse-graining method:

$$x_{cg}^{\tau} = \{\frac{x_1 + x_2 + x_3}{3}, \frac{x_4 + x_5 + x_6}{3}, \dots, \frac{x_{N-2} + x_{N-1} + x_N}{3}\} \tag{7}$$

But for the modified graining method:

$$x_{mg}^{\tau} = \{\frac{x_1}{3}, \frac{x_2 + x_1}{3}, \frac{x_3 + x_2 + x_1}{3}, \frac{x_4 + x_3 + x_2}{3} \dots, \frac{x_N + x_{N-1} + x_{N-2}}{3}\} \tag{8}$$

This method can adapt to various data structures and trends since it averages points in a sliding window manner, allowing for potentially better handling of local variations.

The second step of the MSAE analysis is to calculate the entropy of the modified-grained signal $x_{mg,j}^{\tau}$ for each scale factor τ. Different types of entropy measures can be used in this step. In this study, one of the newest entropies, attention entropy (AttnEn), presented in the previous sub-section, is used because it is a parameter-free entropy.

3. Results

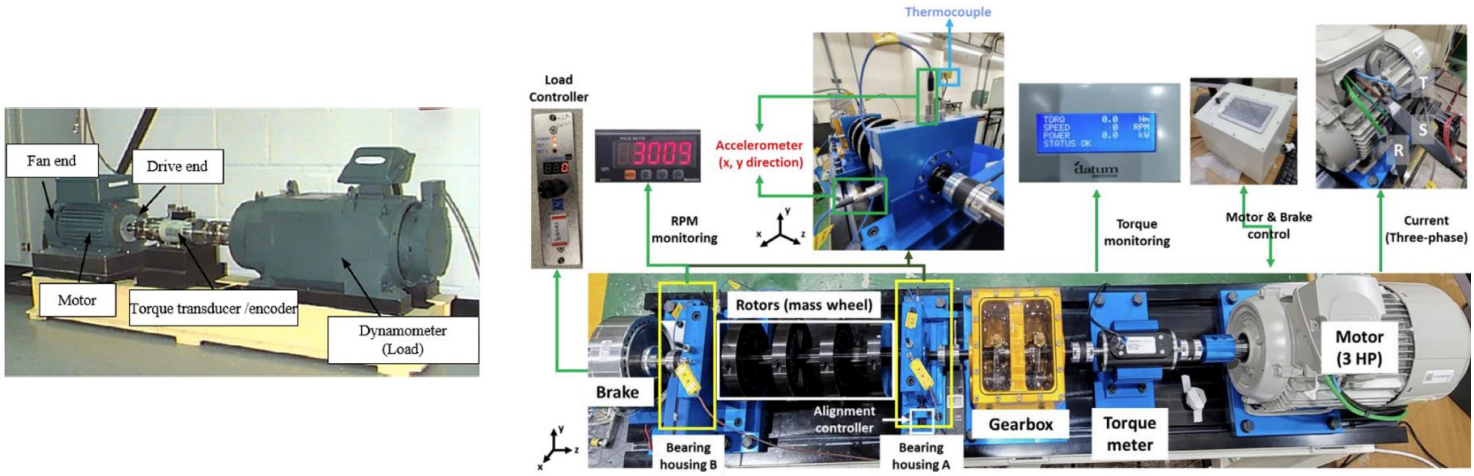
In this section, the effectiveness of the feature vector obtained from the MSAE is compared to feature vectors built from the traditional features on two datasets in two scenarios. The MSAE implementation to a signal results in a τ_{opt}-length feature vector. To have a fair comparison, a τ_{opt}-length feature vector built using the MSSE is considered to compare two multi-scale entropies. Also, eight sets of τ_{opt}-length feature vectors are selected randomly from the 32-features list in Table 1.

In this study, two different datasets are used. The first dataset is the Case Western Reserve University bearing (CWRU) dataset. Vibration signals were collected at 12 kHz or 48 kHz for normal bearings and damaged bearings with single-point defects under four different motor loads from a test rig shown in Figure 2 (a). Within each working condition, single-point faults were introduced with fault diameters of 0.007, 0.014, and 0.021 inches on the rolling element, the inner ring, and the outer ring, respectively. In this paper, we used the data collected from the drive end, and the sampling frequency was equivalent to 12 kHz with no load condition. The second dataset is a more challenging dataset, collected from a rotor-gearbox-bearing test rig in the Korea Advanced Institute of Science and Technology (KAIST), as shown in Figure 2 (b). This dataset was collected using four accelerometers, two thermocouples, three CT sensors, an acoustic microphone, and one tachometer. Also, the main dataset was obtained under different loads and in various parts with or without constant rotating speeds. Three levels of severity for four faults, including shaft unbalance, shaft misalignment, bearing inner ring fault, and bearing outer ring fault are considered in this dataset. However, in our study, we use only vibration signals, collected by four ceramic shear ICP-based accelerometers (PCB352C34) in no load condition and 3010 rpm constant rotating speed. These accelerometers are mounted on x- and y-directions of two-bearing housings based on ISO 10816-1:1995.

It can be seen in Figure 2 that two datasets were acquired from two different test rigs. It leads to different signal distributions and different patterns behind the signals in the two datasets. One of them is only a bearing test rig but another is a rotor-bearing test rig. It is true that in both datasets, there are bearing faults but they are different. In the rotor-bearing systems, a weak fault in another part of the system can have effects on the signals acquired for bearing faults, while in the bearing system, it is only the bearing and faults related to it.

Table 1. A set of 32 statistically and probabilistically features.

	Feature Name		Feature Name		Feature Name		Feature Name
1	Mean	**9**	Skewness	**17**	Signal range	**25**	Zero-crossing
2	Min	**10**	Kurtosis	**18**	Waveform length	**26**	Peak intensity
3	Max	**11**	Absolute Skewness	**19**	Slope sign change	**27**	Signal slope mean
4	Min-max	**12**	Absolute Kurtosis	**20**	Impulse factor	**28**	Mean frequency
5	Standard deviation	**13**	Mean absolute deviation	**21**	Clearance factor	**29**	Max frequency
6	Mean square	**14**	Inter-quartile range	**22**	Shape factor	**30**	Median frequency
7	Root mean square	**15**	Auto-correlation	**23**	Margin factor	**31**	Energy
8	Variance	**16**	Crest factor	**24**	Peak to peak value	**32**	Log sum



(a) (b)

Figure 2. The test rig layouts of the (a) CWRU dataset [9], and (b) the KAIST dataset [10].

Two scenarios are considered in this study. In the first scenario, named fault diagnosis (FD) scenario, the fault severity is not considered and in the second scenario, named fault diagnosis and severity estimation (FD&SE) scenario, the fault severity is considered. The labels used in the CWRU and KAIST datasets, and also the number of classes for two scenarios are mentioned in Table 2.

Table 2. The labels used for the CWRU and the KAIST datasets in the FD and the FD&SE scenarios.

Row	CWRU dataset			KAIST dataset		
	State	FD	FD&SE	State	FD	FD&SE
1	Normal	0	0	Mild misalignment fault	0	0
2	Mild inner ring fault	1	1	Moderate misalignment fault	0	1
3	Moderate inner ring fault	1	2	Severe misalignment fault	0	2
4	Severe inner ring fault	1	3	Mild outer ring fault	1	3
5	Mild ball fault	2	4	Moderate outer ring fault	1	4
6	Moderate ball fault	2	5	Severe outer ring fault	1	5
7	Severe ball fault	2	6	Mild inner ring fault	2	6
8	Mild outer ring fault	3	7	Moderate inner ring fault	2	7
9	Moderate outer ring fault	3	8	Severe inner ring fault	2	8
10	Severe outer ring fault	3	9	Normal	3	9

To obtain the optimal time scale of the MSAE, the test accuracy of the SVM classifier using feature vectors obtained, using different time scales, are investigated on the CWRU and KAIST datasets and only in the FD&SE scenario. For CWRU dataset, the accuracy is increased by increasing

the time scale from 2 to 4, then, the accuracy remained constant and computational costs increased for more time scales (Table 3). Therefore, the optimal time scale, τ_{opt}, is four, and all of the feature vectors must have a length of four. For the KAIST dataset, the same procedure is done. However, the time it takes to obtain an accuracy of about 99%, is 738.93 seconds only for preprocessing of the data, which is very high compared to the time-consumption of the MSAE on the similar time scale on the CWRU dataset. The results are presented in Table 4.

To obtain the accuracy of the presented feature extractor based on the MSAE and compare it with the other ten feature extractors, a comprehensive study is conducted. First, the CWRU and KAIST dataset's signals are divided into 2048 data-points signals, without any overlapping. Then, using the optimal time scales of the MSAE, ten defined 4-length, and 8-length feature vectors are computed for every signal in CWRU and KAIST datasets, respectively. Next, the features are entered into an SVM classifier to diagnose the faults or further estimate the fault severities in two defined scenarios. This procedure is repeated twenty times and the obtained test accuracies, averages, and quartiles are computed. This is done to obtain the feature extractor performance. When the maximum and minimum accuracy of one model during twenty iterations is close to each other, the model presents more robust performance. Also, when the mean of accuracies is high the model is more accurate. The results are illustrated as a box plot diagram in Figure 3 and Figure 4, for FD and FD&SE scenarios, respectively.

Table 3. Comparing test accuracies and time of data pre-processing of the CWRU dataset in the FD&SE scenario using MSAE feature extractor.

Number of scales	Test Accuracy	Times (s)
32	98.52%	95.56
16	98.56%	51.38
8	98.43%	26.69
4	**98.43%**	**14.91**
3	89.92%	10.06
2	71.27%	8.53

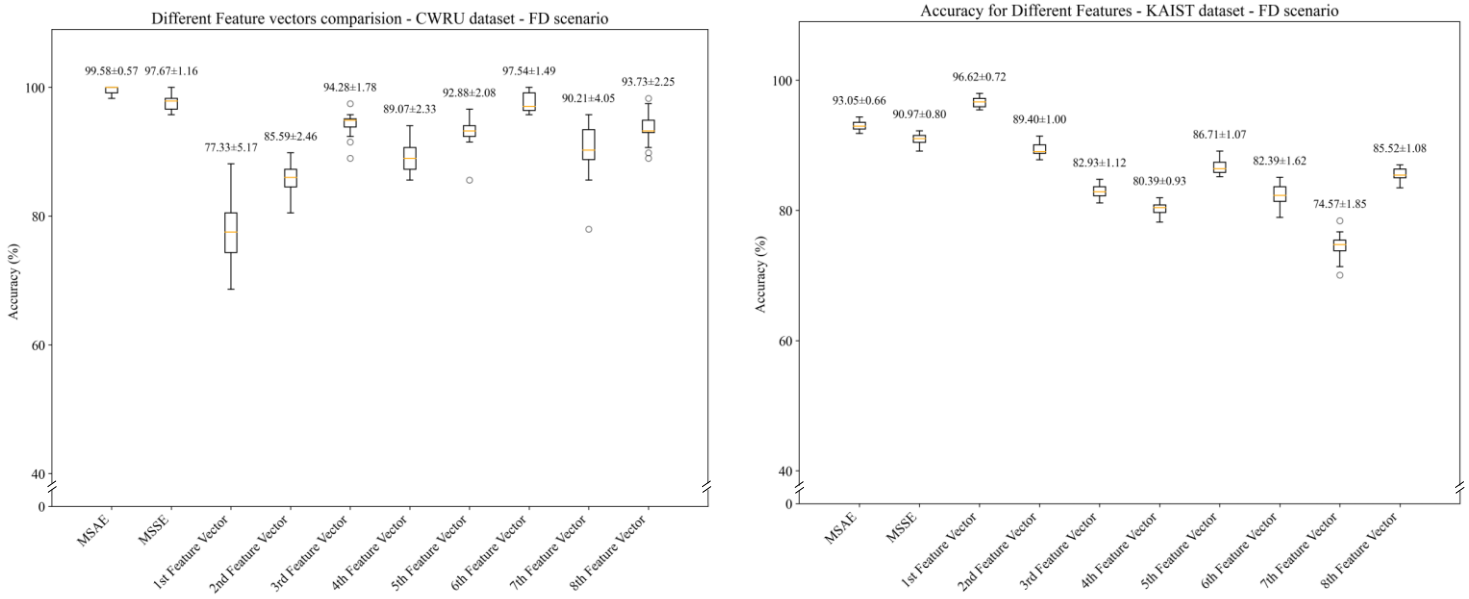
Table 4. Comparing test accuracies and time of data pre-processing of the KAIST dataset in the FD&SE scenario using MSAE feature extractor.

Number of scales	Test Accuracy	Times (s)
32	99.01%	738.93
16	98.43%	372.71
8	**95.89%**	**188.89**
4	87.81%	114.73
3	82.25%	75.55
2	65.02%	52.64

It can be observed that the accuracy of diagnosing the faults is almost larger than estimating their severities as expected. Also, the overall accuracies obtained for the KAIST dataset are lower than those obtained for the CWRU dataset. This is because the KAIST dataset is more challenging. The tolerance of accuracies for 20 iterations, shows the robustness of the MSAE features. For instance, the tolerance for the first feature vector in Figure 3 (a) is 5.17% but for the MSAE feature vector, it is only 0.57%. This tolerance in two scenarios and for both datasets is less than 0.75%.

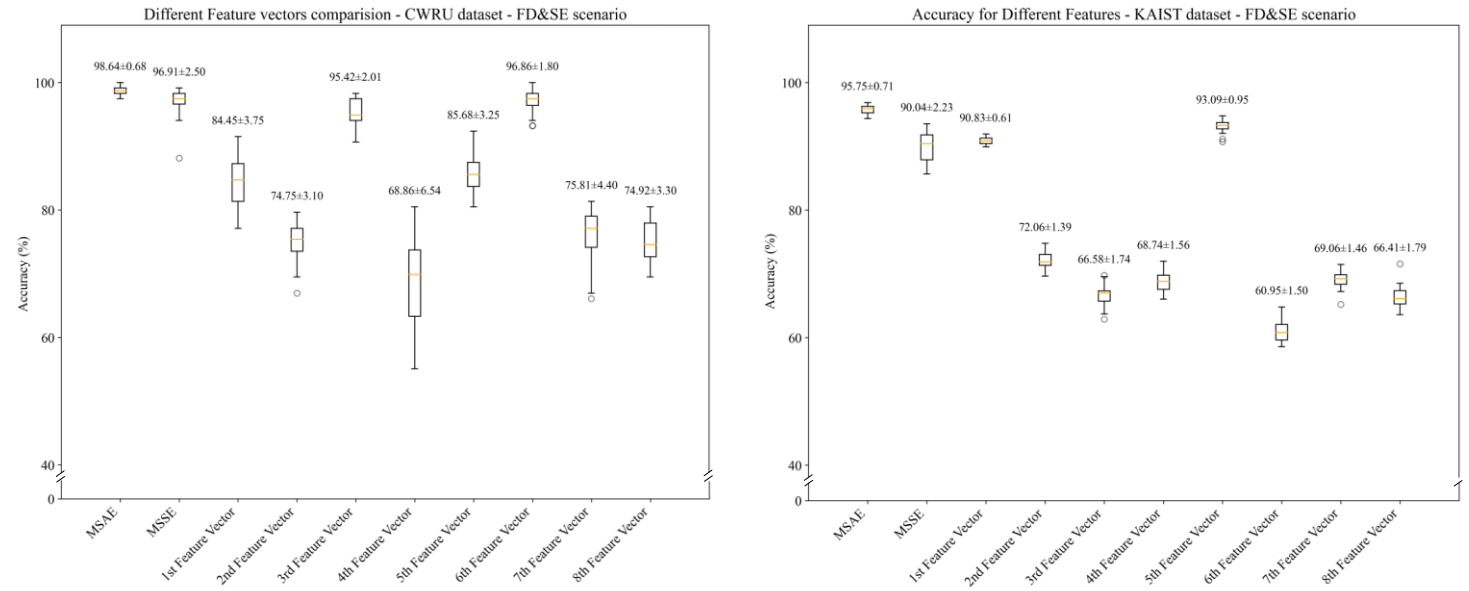
Another important result obtained from seeing the box plots is that the MSAE features provide the most efficient feature set for both scenarios and in both datasets. However, some of the eight random feature vectors can achieve similar accuracies as MSAE, better or worse accuracies. It is worth noting that it is dependent on the selection of the features from the 32 features in Table 1, and

this selection is very dependent on the prior knowledge of the user from the system and the environmental conditions, while the MSAE does not require any prior knowledge.



(a) (b)

Figure 3. The box plot of the FD scenario tests accuracy for the ten different feature vectors using (a) the CWRU dataset and (b) the KAIST dataset.



(a) (b)

Figure 4. The box plot of the FD&SE scenario tests accuracy for the ten different feature vectors using (a) the CWRU dataset and (b) the KAIST dataset.

It is worthy to illustrate the AttnEn values for different time scales in two scenarios. As can be seen in Figure 5, using the trends of changes in AttnEn values for different scales can be more useful than using just the original signal entropy values. It is more obvious for the FD&SE scenario, where the classes are more and the AttnEn values of different classes are very close to those of other classes

but the trends are different. In addition, the trends of changes in AttnEn for the CWRU dataset are different from those for the KAIST dataset. This is due to the differences between the two datasets, which were discussed in the previous sections.

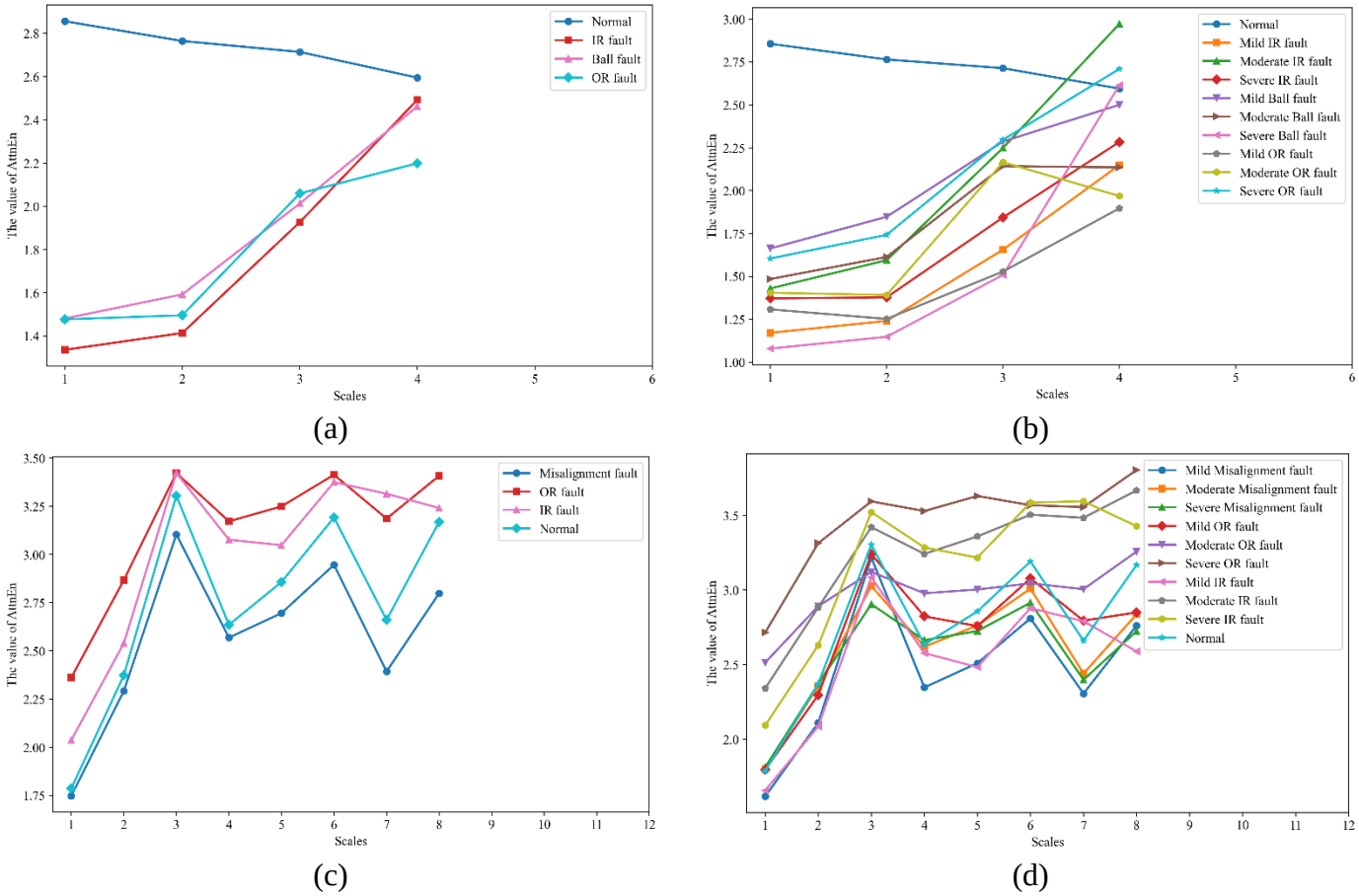


Figure 5. The value of AttnEn for different scales in (a) the FD scenario on the CWRU dataset, (b) the FD&SE scenario on the CWRU dataset, (c) the FD scenario on the KAIST dataset, and (a) the FD&SE scenario on the KAIST dataset.

4. Conclusions

In this study, a new framework was introduced and used for fault diagnosis and fault severity estimation of rotating machines. This framework was made of a robust and free-of-prior-knowledge feature extraction method named MSAE, which is a combination of AttnEn and MSE methods, and an SVM classifier. To investigate the efficiency and performance of this novel framework, the CWRU and the KAIST datasets were used in two scenarios. In the first scenario, the fault diagnosis of rotating machines, and in the second, the severity levels of each fault were considered, respectively. To have a fair comparison, the MSAE method was compared with the earliest version of multi-scale entropies named MSSE beside eight feature vectors built randomly from a set of 32 time-domain well-known statistically and/or probabilistically features. This comparison was based on the test accuracy and was repeated 20 times. According to the time scales of the MSAE and MSSE methods, the feature vectors resulting from them were varied in length. Therefore, the optimal time scale for the MSAE method was obtained differently for the CWRU and KAIST datasets. Briefly, the important results are listed below:

- The fault diagnosis accuracy of the MSAE-SVM model for CWRU and KAIST datasets were $99.58\% \pm 0.57\%$ and $93.05\% \pm 0.66\%$, respectively. These results were superior to all of the other models in this study except for the first random feature vector and KAIST dataset.
- The fault diagnosis and severity estimation accuracy of the MSAE-SVM model for CWRU and KAIST datasets were $98.64\% \pm 0.68\%$ and $97.75\% \pm 0.71\%$, respectively. These results were superior to all of the other models.
- The fluctuation of the averaged testing accuracies is very significant for the eight random feature vectors made up of a set of 32 features. This is because the procedure of selecting the best traditional feature vectors, is very dependent on having prior knowledge of the problem and the inherent physics of the datasets.
- The tolerance of the accuracies corresponding to the feature vectors, randomly selected from a set of 32 time-domain features, is much wider than that of the MSAE feature vector. For example, in the FD&SE scenario and on the CWRU dataset, the tolerance of the 4^{th} feature vector is 6.54% while it is only 0.68% for the MSAE feature vector. Therefore, the MSAE feature vectors present more robust testing accuracies.
- The AttnEn values in different time scales and for various classes showed an interesting trend. One can conclude that using only one AttnEn scale may result in a misunderstanding of the pattern behind the classes and consequently a disappointing classification while using multi-scales simultaneously results in a better classification.

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, pp. 1–15, 2022, doi: 10.3390/s22052012.

[2] Z. Shen, X. Chen, X. Zhang, and Z. He, "A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM," *Measurement*, vol. 45, no. 1, pp. 30–40, 2012, doi: 10.1016/j.measurement.2011.10.008.

[3] C. Y. Yang and T. Y. Wu, "Diagnostics of gear deterioration using EEMD approach and PCA process," *Measurement*, vol. 61, pp. 75–87, 2015.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[5] J. Yang, G. I. Choudhary, S. Rahardja, and P. Franti, "Classification of Interbeat Interval Time-Series Using Attention Entropy," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 321–330, 2023, doi: 10.1109/TAFFC.2020.3031004.

[6] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale Entropy Analysis of Complex Physiologic Time Series," *Phys. Rev. Lett.*, vol. 89, no. 6, 2002, doi: 10.1103/PhysRevLett.89.068102.

[7] Y. Li, S. Wu, J. Chen, and P. Ba, "A fault diagnosis method on GWO-SPA and MAE for fault diagnosis of reciprocating compressors," in *Journal of Physics: Conference Series*, 2023. doi: 10.1088/1742-6596/2528/1/012037.

[8] Z. Liang, "Novel method combining multiscale attention entropy of overnight blood oxygen level and machine learning for easy sleep apnea screening," *Digit. Heal.*, vol. 9, 2023, doi: 10.1177/20552076231211550.

[9] "Case western reserve university (CWRU) bearing data center." Accessed: Dec. 17, 2023. [Online]. Available: https://engineering.case.edu/bearingdatacenter/apparatus-and-procedures

[10] W. Jung, S. H. Kim, S. H. Yun, J. Bae, and Y. H. Park, "Vibration, acoustic, temperature, and motor current dataset of rotating machine under varying operating conditions for fault diagnosis," *Data Br.*, vol. 48, 2023, doi: 10.1016/j.dib.2023.109049.