Proceedings of the Canadian Society for Mechanical Engineering International Congress
32nd Annual Conference of the Computational Fluid Dynamics Society of Canada
Canadian Society of Rheology Symposium
CSME-CFDSC-CSR 2025
May 25–28, 2025, Montréal, Québec, Canada

Bearing Fault Diagnosis Using Text Analysis on the CWRU Dataset

Zehui Hua, Mert Sehri, Patrick Dumond
Department of Mechanical Engineering, University of Ottawa, Ottawa, Canada
*zhua079@uottawa.ca

Abstract— Over the past decade, there has been increasing use of machine learning (ML) algorithms for fault detection by ML researchers. This emergence is due to the ever growing need for diagnosing bearing components to prevent catastrophic machine failures in industry. More recently, large language models (LLMs) have raised huge attention from both researchers and engineers. LLMs could be used in many applications like text classification and sentiment analysis. Compared with the commonly used inputs in intelligent fault diagnosis, fault diagnosis based on text analysis is very intuitive and easy to understand and thus shows great potential. This paper presents a preliminary analysis of bearing data by extracting features into a text file, followed by ML algorithm analysis to achieve accurate diagnostics.

Keywords: machine learning; bearing fault diagnosis; feature extraction; text extraction; CWRU;

I. INTRODUCTION

Diagnosing bearing components has become increasingly critical for preventing catastrophic machine failures, which can lead to substantial financial losses and safety hazards [1]. Bearings cause 50% to 60% of machinery failures in industry [2]. Hence there is a need for ML algorithm analysis for fast and accurate fault detection to prevent machine failures [3].

Manual inspection and basic signal processing methods have traditionally been the cornerstone of fault detection, yet these approaches often fall short in identifying early indicators of bearing fault conditions [4]. Machine learning (ML) algorithms, on the other hand, provide a more advanced alternative, leveraging extensive datasets to uncover patterns indicative of a machine's health [5]. The key to training such machine learning models is to connect the collected vibration signal with its corresponding health state. Once the end-to-end model is well trained, it can accurately predict the actual health states of the input data in a timely manner without human intervention.

Although ML algorithms hold great promise for fault detection, their adoption in industrial settings faces obstacles such as computational demands and low diagnostic accuracy [6]. These issues stem from the complexity of ML models and the substantial amount of labeled data required for training. Furthermore, challenges like fluctuating operational conditions and noise in the data add layers of difficulty to achieving reliable diagnostics [7].

Most current studies focus on preparing datasets by sliding a window in the time domain or in the frequency domain after applying a fast Fourier transform (FFT). Since vibration data is usually collected using a high sampling frequency, a large-sized input signal is obtained, represented as the dimension of the collected signal. Additionally, amplitudes across different working conditions vary significantly. To generate a synthetic dataset with a smaller input size, text provides an alternative solution. This preliminary study delves into bearing data by extracting features and saving them to text files, which are then analyzed using ML algorithms. By transforming raw data into a structured format, the process aims to enhance efficiency and accuracy in detecting faults. The extracted features are evaluated using different ML models to assess their capability in diagnosing bearing-related issues.

This approach seeks to address the limitations of current fault detection methods by reducing computational times and enhancing diagnostic accuracies. Through this preliminary analysis, the aim is to demonstrate the feasibility of using ML algorithms for bearing fault detection in text data applications. The findings of this study will encourage ML researchers to explore these techniques in research, ultimately leading to more reliable and efficient machinery maintenance practices.

The integration of ML algorithms for fault detection holds great promise for improving the reliability and efficiency of machinery maintenance. By overcoming the challenges of high computational times and low diagnostic accuracies, researchers can ensure the continued safe operation of critical machinery components.

II. METHODS

This section expands on the feature extraction methods used for preprocessing raw bearing data. Then a brief introduction of the fault diagnosis algorithms used is given later.

A. Extracted features

1) Mean

Mean denotes the average value of a dataset, calculated by summing all the values and dividing by the number of values, as expressed below:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

where x_i represents the signal from the time domain directly. The total number of x values is n. The final average value is recorded by $\overline{x}$.

2) Standard Deviation (Std)

Std (σ) denotes a measure of the amount of variation or dispersion in a dataset, indicating how spread out the values are from the mean, as formulated by:

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \tag{2}$$

3) Root Mean Sqaure (RMS)

The square root of the average of the squares of a set of values, often used to measure the magnitude of a varying quantity:

$$x_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{3}$$

4) Maximum Absolute Value

The maximum absolute value represents the largest deviation from zero:

$$x_{max} = \max\left(|x|\right) \tag{4}$$

5) Skewness

A measure of the asymmetry of the probability distribution of a dataset, indicating whether the data is skewed to the left or right. The skewness can also be understood as the third-order expectation of x:

$$x_{skew} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3}{\sigma_x^3} \tag{5}$$

6) Kurtosis

A measure of the "tailedness" of the probability distribution of a dataset, indicating the presence of outliers. As local failures related to bearings will cause abnormal impulses, kurtosis is widely used to show how many impulses exist in the objective data. Statistically, kurtosis is the fourth-order expectation:

$$x_{kurt} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4}{\sigma_x^4} \tag{6}$$

7) Mean in the Frequency Domain

Besides the time domain, the frequency domain is also widely used, as fault related frequencies can be calculated by functions once the specification parameters of the bearing are given:

$$\overline{f} = \frac{1}{n}\sum_{i=1}^{n} f_i \tag{7}$$

where f_i denotes the corresponding Fourier spectrum of the time domain signal, obtained by applying the fast Fourier transform, for example.

8) Std in the Frequency Domain

This parameter measures the spread or dispersion of the frequency components around the mean frequency:

$$\sigma_f = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f_i - \overline{f}\right)^2} \tag{8}$$

9) RMS in the Frequency Domain

RMS in the frequency domain expresses how the power of a signal is distributed across the different frequencies:

$$f_{Sqrt} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} f_i^2} \tag{9}$$

B. Fault Diagnosis Algoirthms

Once statistical features are extracted into a text file, machine learning algorithms use this file as an input. Both samples and their corresponding labels (health states) are fed into the model for training, which enforces a supervised learning approach. After training, the model will be able to determine whether the bearing under investigation is healthy or faulty.

1) K-Nearest Neighbor (KNN)

KNN is a simple shallow supervised ML algorithm. As a result, KNN is developed based on the assumption that the training data and testing data share the same feature distribution. Thus, it has limitations when it comes to domain shift (i.e., model performance may degrade under variable working conditions). However, KNN models are very fast to train. Since KNN models can achieve 100 % accuracy on training datasets, overfitting of this data can often occur, which also limits the generalization capacity of the model to new testing data.

2) Convolutional Neural Network (CNN)

To automatically extract features from the originally collected vibration data, deep learning methods such as CNNs are widely used due to their powerful feature representation performance.

The network structure used in this study is described in Table I, where the first 4 layers are convolutional layers and the last two layers are fully connected layers [8]. n_c denotes the number of classes.

The class loss of the CNN during training is expressed as:

$$\mathcal{L}_{class} = \mathbb{E}_{(,)\sim\mathbb{P}^{tr}}\ell\left(y_s, \hat{y}_s\right) \tag{10}$$

$$\hat{y}_s = h_c\left(h_b\left(h_f\left(x_s\right)\right)\right) \tag{11}$$

where (x_s, y_s) represents the sample from the source domain, and h_c, h_b, h_f denote the classifier, bottleneck, and feature extractor, respectively.

TABLE I. NETWORK STRUCTURE OF THE CNN FOR FEATURE EXTRACTION

Layers	***Modules***
Layer 1	Conv1(1, 16, 5), BN(16), ReLU
Layer 2	Conv1(16, 32, 3), BN(32), ReLU, Maxpool(2)
Layer 3	Conv1(32, 64, 3), BN(64), ReLU
Layer 4	Conv1(64, 128, 3), BN(128), ReLU, AdaptiveMaxpool(4)
Layer 5	Linear(512, 256), ReLU, Linear(256, 256), ReLU
Layer 6	Linear(256, n_c)

By implementing different machine learning algorithms, a comparison study can be conducted to evaluate their performance on the same dataset. It is worth noting that all these methods could provide an end-to-end diagnosis based on the given data so that human intervention can be avoided.

III. RESULTS AND DISCUSSION

A. The CWRU Bearing Dataset

As a widely used bearing dataset, Case Western Reserve University (CWRU) provides 10 classes of bearing data that are collected under 4 different load conditions [9]. The experimental test rig is shown in Figure 1.

Figure 1. CWRU bearing test rig [9].

The applied loads vary from 0 hp to 3 hp. The rotation speed is set as constant. The 10 different classes include normal (N), inner race fault (I), outer race fault (O) and ball fault (B) combined with three different fault sizes.

The domains built using the CWRU bearing dataset are listed in Table II. Domains are constructed based on the 4 different load conditions. A window length of 1024 is selected to generate each sample from the original data, and the overlap between two sequential samples is set as 256 (25 %). As a result, 1000 samples (100 samples in each class and 10 classes) with a length of 1024 are generated.

TABLE II. DOMAINS BUILT USING THE CWRU BEARING DATASET

Domains	***Settings***		
	Loads	***Health States***	***Sample Numbers***
1	0 hp	N, I007, I014, I021 O007, O014, O021, B007, B014, B021	10×100×1024
2	1 hp		10×100×1024
3	2 hp		10×100×1024
4	3 hp		10×100×1024

B. Text Generation and Tokenization

Once the dataset is truncated in the time domain is generated, preprocessing is then used to transform the features into sentences, as sentences can be easily understood by researchers and users, which is believed to help improve the readability of the collected data. The flowchart in Figure 2. illustrates how text is generated from the bearing data. After converting the selected statistical features into texts in a structured format, the readability could be further improved.

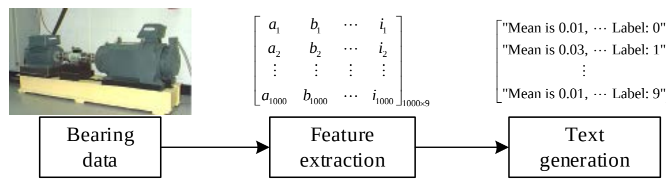

Figure 2. Dataset preparation on the CWRU bearing data.

As for tokenization, this technique is applied to transform the generated texts into a series of numbers again, by referring to a dictionary [10]. Each word/figure in the generated text files will be substituted with an index consisting of numbers. This enables the training model to understand the input based on the newly constructed dictionary with different indices, which in turn shortens the input size of the training data. Feature visualization results of the texts after tokenization are given in Figure 3.

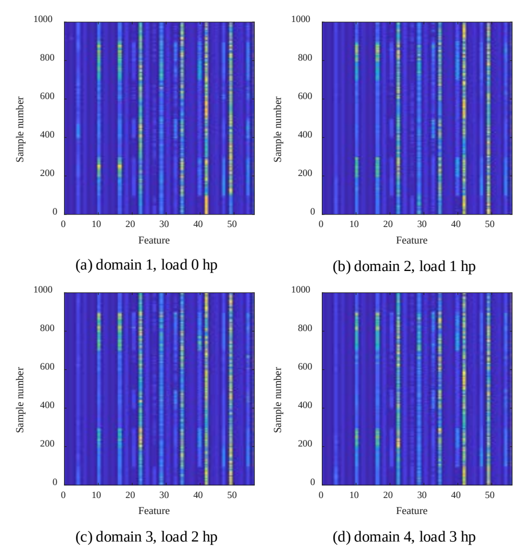

Figure 3. Feature visualization of the texts after tokenization.

Finally, the original input of size [1000×1024] in the time domain is transformed to a new size of [1000×56]. Horizontally, the plotted features in different colors denote the indices referring to the newly constructed dictionary after tokenization, which have a length of 56. Vertically, the y axis denotes the number of samples, and a total of 1000 samples are drawn because there are 10 classes, and 100 samples are generated in each class.

C. Text ML Results

Once all sentences are generated, the model can be trained. Then two different ML methods are tested, including KNN and CNN.

As a traditional supervised ML method, KNN can be used to classify data based on the similarity of nearest neighbors. On the other hand, CNNs are neural networks with specific layers, that extend their scope to deep learning. CNNs are designed to automatically extract features through convolutional operations. The results obtained on the four different domains are listed in Table III. It can be found that the CNN outperforms KNN in acquiring a higher testing accuracy, whereas KNN is a supervised ML algorithm, which leads to 100 % accuracy on training data, the testing accuracy drops significantly compared with the CNN as testing samples are not involved during training.

TABLE III. ACCURACY RESULTS ON THE CWRU BEARING DATASET WITHIN A SINGLE DOMAIN (%)

Methods	*Domains*	*Training Accuracy*	*Testing Accuracy*
KNN	1	100	72.50
	2	100	66.00
	3	100	64.50
	4	100	62.00
CNN	1	98.62	85.97
	2	95.25	87.96
	3	95.00	85.46
	4	97.88	91.46

The training and testing accuracy when training the CNN are plotted in Figure 4. Loss versus epoch is also drawn to show that after a certain number of epochs (i.e., 80 epochs), the performance of the model converges. By comparing the training and test accuracy, a slight difference can be observed since no labeled test samples contribute to the model training procedure.

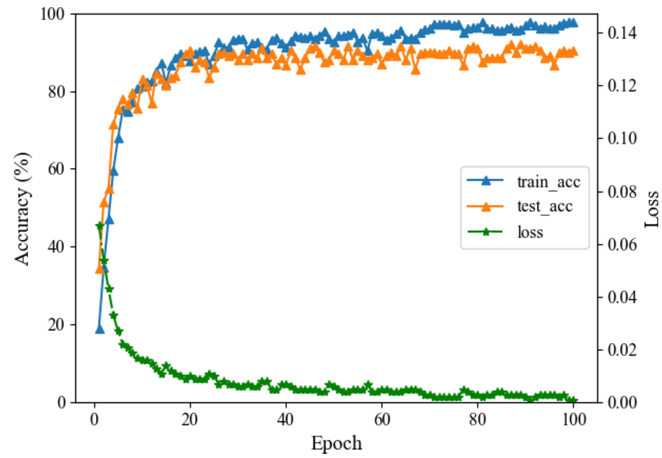

Figure 4. Accuracy and loss versus epochs on domain 4 (load 3 hp).

Also, transfer tasks that are built between two different domains are tested. For instance, task 1-2 denote that the model is trained using domain 1 and tested using domain 2. The results are listed in Table IV. It can be found that CNNs still outperform KNN when looking at the generalization capacity of the sequential data. The accuracy obtained by using a CNN varies from 62 % to 76 %, showing that these two domains share some similarities.

TABLE IV. ACCURACY RESULTS ON THE CWRU BEARING DATASET BETWEEN DIFFERENT DOMAINS (%)

Methods	*Tasks*	*Target Training Accuracy*	*Target Testing Accuracy*
KNN	1-2	43.75	40.00
	1-3	35.00	34.00
	1-4	39.63	38.00
	2-3	57.75	61.00
	2-4	53.00	52.50
	3-4	48.00	48.50
CNN	1-2	62.62	69.44
	1-3	62.88	68.35
	1-4	62.88	62.34
	2-3	74.38	76.92
	2-4	79.25	70.88
	3-4	71.50	71.93

IV. CONCLUSION

This paper explores the application of a text analysis technique for fault diagnosis. The methodology involves the manual extraction of 9 distinct statistical features from both the time and frequency domains. These features are then used to generate textual representations of the data. To facilitate computational understanding, a Tokenizer is employed to convert the original text into a sequential data format.

This study utilizes two ML models, KNN and a CNN, to evaluate the effectiveness of the proposed text analysis-based bearing fault diagnosis method. The results on the CWRU dataset show that text analysis-based fault diagnosis can provide an accuracy of nearly 90%, indicating that the use of text files instead of commonly used time series data can provide an alternative dataset preparation method for bearing fault diagnosis. Code used in this paper is available at:

https://github.com/jshzh163com/Text_analysis_for_fault_diagnosis

REFERENCES

[1] R. Collacott, *Mechanical Fault Diagnosis and condition monitoring*. Springer Science & Business Media, 2012.

[2] M. Sehri, P. Dumond, and M. Bouchard, "Design of a Data Fusion and Deep Transfer Learning Test Rig for Roller Bearings Diagnosis and Prognosis," presented at the ASME 2023 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection, Nov. 2023. doi: 10.1115/DETC2023-117264.

[3] G. Vashishtha *et al.*, "Advancing machine fault diagnosis: a detailed examination of convolutional neural networks," *Meas. Sci. Technol.*, vol. 36, no. 2, p. 022001, Dec. 2024, doi: 10.1088/1361-6501/ada178.

[4] F. Alharbi *et al.*, "A Brief Review of Acoustic and Vibration Signal-Based Fault Detection for Belt Conveyor Idlers Using Machine Learning Models," *Sensors (Basel)*, vol. 23, no. 4, p. 1902, Feb. 2023, doi: 10.3390/s23041902.

[5] L. Zhang, J. Lin, B. Liu, Z. Zhang, X. Yan, and M. Wei, "A Review on Deep Learning Applications in Prognostics and Health Management," *IEEE Access*, vol. 7, pp. 162415–162438, 2019, doi: 10.1109/ACCESS.2019.2950985.

[6] D. Leite, E. Andrade, D. Rativa, and A. M. A. Maciel, "Fault Detection and Diagnosis in Industry 4.0: A Review on Challenges and Opportunities," *Sensors (Basel)*, vol. 25, no. 1, p. 60, Dec. 2024, doi: 10.3390/s25010060.

[7] H. Su and J. Lee, "Machine Learning Approaches for Diagnostics and Prognostics of Industrial Systems Using Open Source Data from PHM Data Challenges: A Review," Sep. 18, 2024, *arXiv*: arXiv:2312.16810. doi: 10.48550/arXiv.2312.16810.

[8] Z. Zhao *et al.*, "Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–28, 2021, doi: 10.1109/TIM.2021.3116309.

[9] "Case Western Reserve University Bearing Data Center Website." [Online]. Available: https://engineering.case.edu/bearingdatacenter.

[10] *Tokenizer: A tokenizer for Icelandic text*. Python. Available: https://pypi.org/project/tokenizer.