Phuong X. NGUYEN, Linh H. TRAN*

Hanoi University of Science and Technology, School of Electrical and Electronics Engineering

doi: 15199/48.2024.05.50

A model for processing and identifying engine vibration signals

Abstract. *The advancement of the sensor technology becoming increasingly cost-effective and the progress in diagnostic and management research, users nowadays not only demand high reliability from their devices but also the ability for their equipment to self-diagnose errors and provide alerts. These devices often incorporate sensor systems capable of generating plenty of data points, that needed a carefully targeted algorithms for extracting features from the data for classification and prediction models. In this paper, we will develop a comprehensive model for identifying vibration signals. We will extract features from the bearing data provided by Case Western Reserve University Bearing Data Center, then use a deep-learning based convolutional neural network to learn to be a classification model of the motor states based on the vibration signals. The numerical results show that the method can offer the promising accuracy at 85.8%.*

Streszczenie. *Wraz z post˛epem w technologii czujników, która staje si˛e coraz bardziej ta´nsza do u˙zycia w badaniach diagnostycznych, u˙zytkownicy wymagaj˛a obecnie nie tylko wysokiej niezawodno´sci swoich urz˛adze´n, ale tak˙ze zdolno´sci ich sprz˛etu do samodiagnostyki bł˛edów i generowania alertów. Nowej generacje urz˛adze´n zawieraj˛a systemy czujników zdolne do generowania mnóstwo danych, co wymagało starannie dobranych algorytmów do wyodr˛ebniania cech charakterystycznych na potrzeby modeli klasyfikacji i predykcji. W tym artykule przedstawimy model do identyfikacji sygnałów drganiowych. Korzystali´smy z danych pomiarowych ło˙zysk dost˛epnych w Centrum baz danych ło˙zysk Uniwersytetu Case Western Reserve. Z tych danych pomiarowych, wygenerowano ich spectrogramy do postaci obrazów a nast˛epnie wykorzystano splotow˛a sie´c neuronow˛a opart˛a na gł˛ebokim uczeniu si˛e do tworzenia model klasyfikacji stanów silnika w oparciu o sygnały wibracyjne. Wyniki liczbowe pokazuj˛a, ˙ze metoda ta mo˙ze zapewni´c obiecuj˛ac˛a dokładno´s´c na poziomie 85,8%. (**Model przetwarzania sygnału drgań silnika**)*

Keywords: vibration signals, fault diagnosis, classification model, feature extraction, deep learning.
Słowa kluczowe: sygnaly wibracyjne, diagnostyka usterek, model klasyfikacji, ekstrakcja cech, głębokie uczenie.

Introduction

In both industry and civil devices, among the compo nents of many kinds of machines, the bearings undoubtedly are one of the most important parts. They have a significant impact on the machinery's performance and reliability. Bear ing failure is a common cause of machinery breakdowns. Consequently, diagnosing bearing faults has emerged as a important research avenue within the research of equipment failure detection.

Utilization of vibration signals to identify faults in the ro tating machines based from the core concept that all electro mechanical apparatus generates distinctive vibrations dur ing their operation. Deviations in electrical circuits and al terations in mechanical structures induce irregular patterns within the vibration spectrum, discernible by monitoring sys tems. Vibration motions propagate across the components, allowing internal malfunctions to contribute their signal pat terns in exterior vibrations, rendering them noticeable to classification and detection systems. This strategy remains equally efficacious even when the machinery is intercon nected, such as in standard production line layouts. When contrasting with alternative sensor types, the merits of vibra tion sensors include:

- Less affected by common industrial noises when com pared to popular audio sensors and/or infrared light sen sors.
- Compact sensor sizes allow for easy installation within existing machines' installation.
- Vibration signals propagate across machines, encom passing information from all components, enabling eas ier detection of internal faults.

Nonetheless, there are still challenges when working with vi bration signals:

- Vibration signals from machines are heavily influenced by both machine and sensor installation.
- Anyalterations to the production lines (e.g., product or technology changes) can impact patter ns and disrupt detection algorithms. . .

These challenges make the application of trainable models for detection and classification of fault states being highly recommended. In this paper we will propose an approach to use the popularCNN (Convolution Neural Network) to pro cess the data from machines' vibration signals to detect the faulty states of the machines. The data samples come from an open-source database available at CWRU [1].

A classification model is typically carried out by a chain of multiple steps, which can include:

- *Signal instrumentation and measurement block*: used for collecting measurable signals from the object to be classified. The vibration signals in this paper was collected using accelerometer sensors [1].
- *Signal pre-processing block*: used to roughly filter noises, and to convert and/or normalize the signals collected from the object. There are numerous factors that can cause various noise components within the system, whether internally within the equipment block or during the data acquisition phase.
- *Feature Extraction Block*: This is the most challenging block in the solution development process. The feature extraction process aims to eliminate undesired feature values and retain high-quality characteristics contributing to accurate classification. In this paper, the spectrogram of the vibration signals was selected to be the main features source of the motors.
- *Classification block*: This block (also known as the recognition model) is responsible for mapping feature vectors to classification results. This mapping can be performed using various methods, corresponding to different recognition models. A trainable CNN model was proposed to recognize the above mention spectrogram of the signals to classify the state of the motors responsible for the vibrations.

The Spectrogram of a time series signal

In this paper, considering that the input signals are vibration signals measured by sensors, they essentially consist of sequences of time-varying values. According to studies in vibration recognition, one fundamental characteristic for identification is the frequency spectrum of the vibration signal. In this model, we use Short-Time Fourier Transform (STFT) to create an "image-like" spectrogram, which is then processed using the CNN model.

Short-time Fourier transform (STFT) is a sequence of Fourier transforms of a windowed signal. STFT provides the time-localized frequency information for situations in which frequency components of a signal vary over time, whereas the standard Fourier transform provides the frequency information averaged over the entire signal time interval [4].

The STFT results as a 2D map of time and frequency being two coefficients are given by:

$$(1)\quad \begin{cases} \chi_{\text{STFT}}[m,n] = \sum_{k=0}^{L-1} x[k]g[k-m]e^{-j2\pi nk/L} \\ x[k] = \sum_m \sum_n X_{\text{STFT}}[m,n]g[k-m]e^{j2\pi nk/L} \end{cases}$$

where $x[k]$ denotes a signal and $g[k]$ denotes an L-point window function. From (1), the STFT of $x[k]$ can be interpreted as the Fourier transform of the product $x[k]g[k-m]$.

The spectrogram is known to be a valuable instrument for examining how the frequency characteristics of a signal evolve in time [3]. The fundamental idea of the spectrogram involves converting time-windowed data from the time domain into the frequency domain to create a two-dimensional "image" of a signal. One dimension shown as one axis indicates the time, the other indicates the frequency, and the color or intensity indicates the signal's magnitude at each point in both time and frequency. Thus, the spectrogram can be interpreted as an illustration of the signal's frequency appearance at a particular time instant. This "image" offers a more comprehensive understanding of the signal compared to analyzing its time-domain waveform or frequency-domain spectrum individually. To demonstrate the usefulness of spectrogram in representing the spectra of signals, let's consider the following example. In this example, a signal of 10 seconds length is the superposition of 3 main components: 50 Hz in the time range $[0s, 4s]$, 80 Hz in the time range of $[4s, 8s]$ and 130 Hz in the time range of $[6s, 10s]$:

(2)
$$\begin{cases} x_1(t) = 10\sin(2\pi \cdot 50 \cdot t) & for \quad 0 \le t \le 4s \\ x_2(t) = 10\sin(2\pi \cdot 80 \cdot t) & when \quad 4 \le t \le 8s \\ x_3(t) = 10\sin(2\pi \cdot 130 \cdot t) & when \quad 6s \le t \end{cases}$$

The spectrogram of the signal $x(t) = x_1(t) + x_2(t) + x_3(t)$ is shown in Fig. 1. It can be seen from the spectrogram that it clearly shows both frequency and time appearance characteristics correctly.

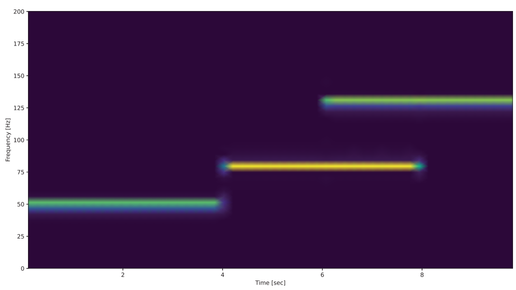

Fig. 1. The spectrogram of the signal defined in (2).

To analyze and recognize the "images" formed from the time-frequency spectra, this paper proposes the use of a customized Convolutional Neural Network (CNN). CNN is a widely adopted tool, extensively used in image recognition and analysis tasks. CNN is relatively simple compared to other new image recognition tools, applicable in numerous domains like facial and object recognition, document analysis, and more. It can also be trained on relatively low-performance computers. The following sections will provide a more detailed explanation of CNN's architecture and learning algorithms.

The Convolutional Neural Network

The Convolutional Neural Network (CNN) is one of the most popular deep learning methods. The Input Layer of CNN is the layer for raw input data, which could be images, audio, or other types of structured data. It's the entry point for the information to be processed by the network. For images, it's treated as a 2D map, where each pixel's values are treated as an input. The CNN structure consists of three main types of layers, each with a distinct role: Convolution layer, Pooling layer, Fully Connected layer [2]. These layers are briefly described as follow.

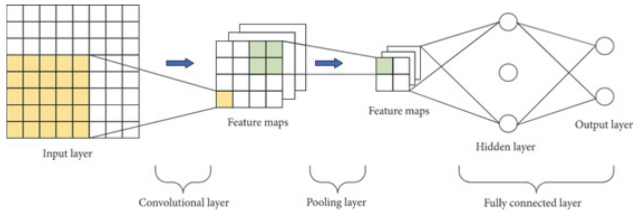

Fig. 2. A typical structure of the CNN network [2].

- The Convolutional layer: This layer is a foundational component within CNN architecture, assuming a pivotal role in the extraction of features from input data, particularly images. In this layer, convolution operations are applied to the input using a set of filters that can be trained. These filters are designed to capture varied features within the data, such as edges, textures, and intricate patterns. The outcome is an assemblage of feature maps that indicate the presence of these diverse features. When repeated, this type of layer can learn more progressive abstract and intricate features. The convolutional layer's ability to automatically detect and extract features makes it a powerful tool for tasks like image recognition and object detection. A convolution layer is denoted by a triple $(N \times W \times H)$, where N is number of filters acting in parallel way, W and H are the width and height of each filter.
- Pooling Layer: This layer play a role in downsampling the spatial dimensions of the feature maps produced by the previous convolutional layers. This helps reduce the computational load and the number of parameters in the network, while retaining important information. A pooling layer is denoted also by a triple $(N \times W \times H)$, where N is number of filters acting in parallel way, W and H are the width and height of each filter. There are two most popular types of pooling, which are called MaxPool and AvgPool, where the output is the maximum or the average of the values being processed by the layer correspondingly. In this paper, the MaxPool layers are used.
- Fully Connected layer: This layer connects every neuron at the input side to each neuron at the output side. It can be used for combining the high-level features learned in earlier layers to make final predictions. For example, in image classification, the Fully Connected layer takes the flattened feature vectors from the previous layers and transforms them into the appropriate output format. The output format depends on the task: for classification, it might be a set of class probabilities; for regression, it could be a single numerical value. The weights and biases in the fully connected layer are adjusted during training to find the best combination of features for accurate predictions.

With those fundamental layers, a CNN can be any combination of layers with no hard limitation on the number of layers nor on the order of the layers. By using *trial-and-error* method, the CNN architecture in this paper was selected with 12 layers (listed with their output shapes) as: Input[96 × 96]-Conv[32 × 3 x 3]-Conv[32 × 3 × 3]- MaxPool[32 x 2 × 2]-Conv[64 × 3 × 3]- Conv[64 × 3 × 3]- MaxPool[64 x 2 × 2]-Conv[128 × 3 × 3]-Conv[128 × 3 × 3]-MaxPool[128 x 2 × 2]-FC[100 x 1]-FC[100 x 1]

Network Parameters Training with Adam Optimizer

The proposed CNN was trained in supervised mode using the Adam optimizer. The optimizer is known for effectiveness in training neural networks by enabling dynamical adaption of the learning rates for different model parameters, making it suitable for various deep learning tasks. The idea of Adam optimizer is to use a combination of momentum and adaptive learning rates to converge to the minimum of the cost function more efficiently [8]. During the training process, it uses the actual mean and variance of the gradients to update the learning rate on the fly.

In this paper, the Adam optimizer used the learning rate $\alpha = 10^{-4}$ and $\epsilon = 10^{-5}$ respectively [8].

Case Western Reserve University Bearing Data Center

The data used in this paper was provided by the Bearing Data Center at Case Western Reserve University [1]. The experiments involved a 2-HP reliance electric motor (left), a torque transducer/encoder (center), a dynamo-meter (right), and accelerometer data were collected from both the drive end and fan end bearings, as illustrated in the Fig. 3. These bearings support the motor shaft.

The faults in the motors' bearing were intentionally introduced into the inner raceway (IR), outer raceway (OR), and rolling element (ball) (BF) of the bearings using electro-discharge machining (EDM), with variations in fault diameter and outer raceway location. The data encompasses motor loads ranging from 0 to 3 HP. The measurements unit for vibration signals based on accelerometers were affixed to the housing at the 12 o'clock position [1]. Each variation were tested and recorded into records of length about 10 seconds each.

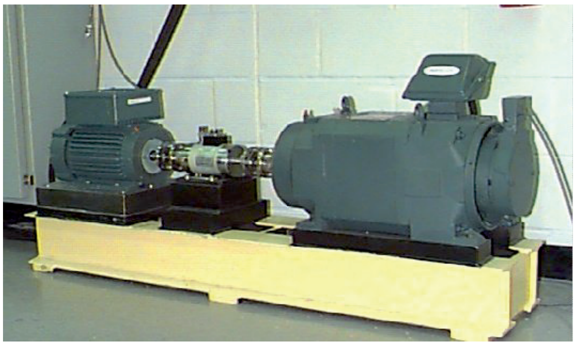

Fig. 3. Test stand for bearing roller for measuring vibration data using accelerometer [1].

The Proposed Model for Processing and Identifying Engine Vibration Signals

The records from CWRU were sampled at 12 kHz and split into segments of 1000 samples to increase the number of samples data generated. The spectrogram of each segment was created using the STFT algorithm with Tukey window [5] with shape parameter of 0.25 [1]. The spectrograms were later resized to be images of 96 x 96 pixels for compatibility of selected CNN network. A total of 49 records for 4 types of motors' state (1 normal and 3 faulty) were used to generated 5853 pairs of input-output samples.

Experimental Results

To prepare the "image" data for the CNN model, we create images using random 4100 (about 70%) of the data pairs for the model training process. The remaining 1753 (about 30%) of the data pairs will be used to generate "images" for the testing process. Examples of raw vibration signals are illustrated in Fig. 4.

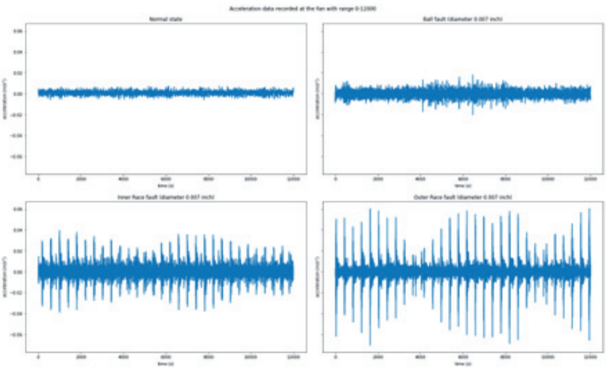

Fig. 4. Example of raw data recorded in CWRU database.

In Fig. 5 and Fig. 6 are shown the generated spectrograms of selected records with different states of the motor. It can be seen that in normal state, the vibration signal contain less of harmonics than in faulty states. Among the faulty states, the fault in rolling balls seem to be more "blurred", indicating that the relative position of the crack is not fixed making the vibration pattern also changed.

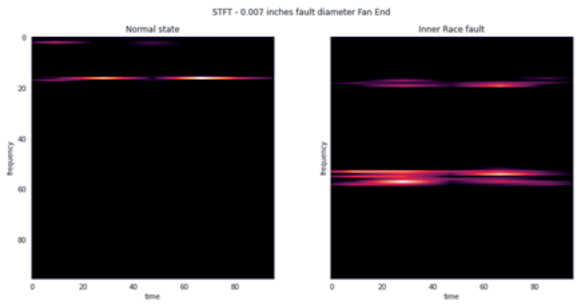

Fig. 5. The data in Normal state and Inner Race fault state shown as their STFT.

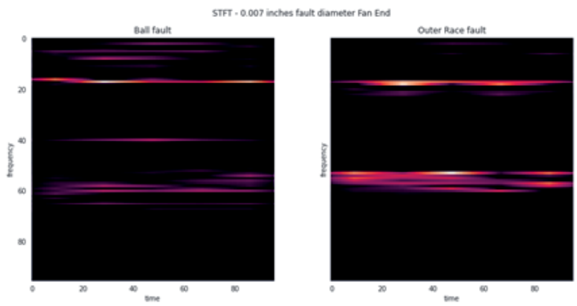

Fig. 6. The data in Ball fault and Outer Race fault states shown as their STFT.

The selected CNN was trained with the Adam optimizer on the generated data sets. The performance of the proposed CNN model is evaluated with the *Receiver Operating Characteristic* (ROC) curve shown in Fig. 7. Based on these ROC curves, it can be concluded that the proposed model has a good discriminatory ability. On the other hand, the value ranging from 0.71 to 0.98, assesses the model's discrimination, with values closer to 0.98 indicating very high discrimination and those closer to 0.71 signifying good but slightly lower discrimination.

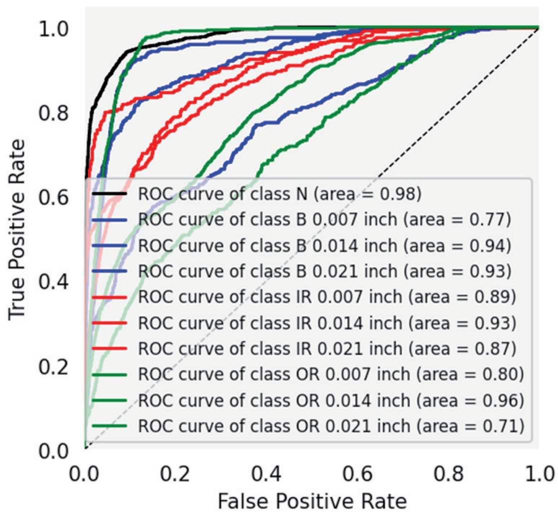

Fig. 7. The Receiver Operating Characteristic curve of the proposed CNN model.

For further analysis, consider the ROC value of class OR 0.021 inch, which is 0.71, the highest among all classes. This suggests that signals from this class have a higher likelihood of being falsely recognized compared to signals from other classes.

The proposed CNN model has achieved results demonstrated in Fig. 8 with the accuracy of 85.8%. From this result, we can say that, overall, the architecture is acceptable but still need some improvement. For instance, according to the Fig. 8, there are defects in recognizing in signal in OR 0.021 inch and the predicted values are significantly inaccurate, a total of 149 samples (adding the numbers in the column 10 except the last value) were misclassified by the classifier, which is the highest misclassification rate among all the classes, the accuracy in prediction for this class is very low, only 30.6% and most of the value is wrongly classed with OR 0.007 inch. This mistake mostly come from the similar in input "images" create by STFT which is shown in the Fig. 9. This problem can potentially be solved by applying the different feature extraction to capture the minor differences between the signals with the same characteristics or recustom the loss functions that penalize misclassifications between the similar signals more heavily.

The current study represents an important step towards more accurate fault diagnosis using the proposed CNN architecture. While the results obtained are promising, it's important to note that further research and refinement are expected to enhance the performance of the model. As technology advances and more data becomes available, the model can be fine-tuned, and its capabilities can be extended. Continued research in this area holds the potential to uncover new insights and techniques that will lead to even more robust and accurate fault diagnosis systems in the future.

Conclusion

This paper introduced a CNN architecture that using time-frequency and image representations of raw vibration signals as an input image for classifying and diagnosing the faults in rolling element bearings. The proposed CNN architecture exhibits an expected good quality result at 85.8% of accuracy and holded a potential for future enhancements. The model can be extended to train with different features extracted from the vibration signals to achieve better accuracy. Or it can be combined with other classification models to improve the reliability of the whole system.

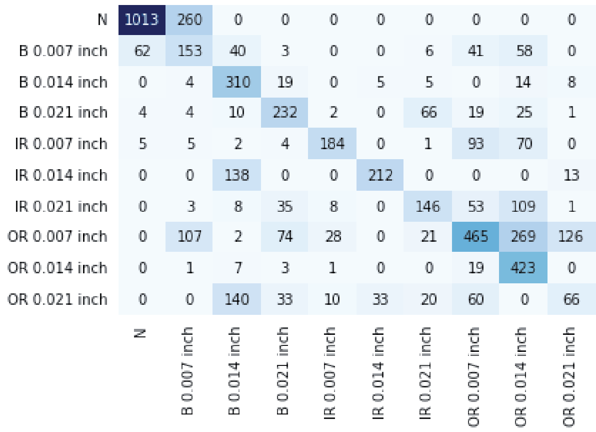

Fig. 8. Confusion matrix for CWRU data for the proposed CNN model.

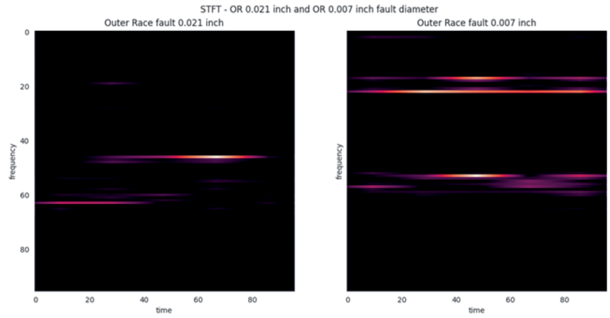

Fig. 9. The sample STFT - OR 0.021 inch and OR 0.007 inch fault.

Authors: *M.Sc. Phuong Nguyen Xuan, School of Electrical and Electronics Engineering, Hanoi University of Science and Technology, email: phuong.nx222187m@sis.hust.edu.vn,*
Assoc. Prof. Dr. Sc. Linh Tran Hoai (Corresponding Author), School of Electrical and Electronics Engineering, Hanoi University of Science and Technology, email: linh.tranhoai@hust.edu.vn.

REFERENCES

[1] Case Western Reserve University Bearing Data Center Website. [web page] `https://engineering.case.edu/bearingdatacenter/` [Accessed on 01 Apr. 2022.].
[2] Venkatesan, Ragav; Li, Baoxin: Convolutional Neural Networks in Visual Computing: A Concise Guide, CRC Press, ISBN 978-1-351-65032-8, 2017.
[3] David Verstraete, Andrés Ferrada, Enrique López Droguett, Viviana Meruane, Mohammad Modarres: Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings, Shock and Vibration, vol. 2017, Article ID 5067651, 2017.
[4] Kehtarnavaz, N.: Frequency Domain Processing, Digital Signal Processing System Design, doi:10.1016/b978-0-12-374490-6.00007-6, pp. 175–196, 2008.
[5] Bloomfield, P.: Fourier Analysis of Time Series: An Introduction, New York: Wiley-Interscience, 2000.
[6] Hua Su, Kil To Chong, R. Ravi Kumar: Vibration signal analysis for electrical fault detection of induction machine using neural networks, Neural Computing and Applications, vol, 20, pp. 183–194, 2011.
[7] Maruthi G. S., Vishwanath Hegde: Application of MEMS Accelerometer for Detection and Diagnosis of Multiple Faults in the Roller Element Bearings of Three Phase Induction Motor, IEEE Sensors Journal, vol. 16, no. 1, pp. 145 – 152, 2016.
[8] Diederik P. Kingma, Jimmy Ba: Adam: A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980, pp. 273 – 297, 2014.