post147

Bearing Health Diagnosis of Rotating Machinery using Deep Transfer Learning Model with Vibration and Audio Signals Input

Wei-Lun Tsao, Ming-Jong Tsai* and Wan-Ru Wu

Graduate Institute of Automation and Control,

National Taiwan University of Science and Technology

Taipei, Taiwan

SCA/HPCAsia 2026

Everything with HPC - AI, Cloud, QC and Future Society

The Supercomputing Asia & The International Conference on High Performance Computing in Asia-Pacific Region

Osaka, Japan, Jan. 26-29, 2026

Abstract

This study proposes a deep transfer learning model integrating vibration and acoustic signals for bearing fault diagnosis of rotating machinery. An experimental system was built using an AC servo motor driven rotating machinery with changeable load, vibration sensor, and directional microphone, and ten replaceable bearings for experiments. Ten bearing modules with different health conditions include normal condition and nine different faults which include three inner ring defects, three outer ring defects, and three ball defects in bearings. Both vibration and acoustic signals were collected and transformed into time-domain and time-frequency images to form three-channel inputs. The VGG19-based model is employed for transfer learning model training, validation, and testing. Transfer learning across different loads (0, 0.5,1 hp) was applied. The experimental results showed that the accuracy of 97.16% (vibration), 98.88% (vibration + sound), and 99.23% (three-channel) are achieved. It demonstrates the effectiveness of the proposed approach for health diagnosis of rotating machinery.

Keywords—Rotating Machine, Transfer Learning, Deep Learning, Bearing Health Diagnosis, VGG19

Technical background

- The Short-Time Fourier Transform (STFT) captures both time and frequency characteristics by applying a window function (e.g., Hanning) to signal segments, as calculated in Equation (1). This process converts 1D signals into 2D time-frequency spectrograms, where the horizontal and vertical axes represent time and frequency, respectively, while color intensity reflects energy levels. The overall transformation process is illustrated in Fig.1.

$$STFT\{x_1[n]\}(m,\omega) \equiv X(m,\omega) = \sum_{n=0}^{N-1} x_1[n]\omega[n-m]e^{(-i\omega n)} \qquad (1)$$

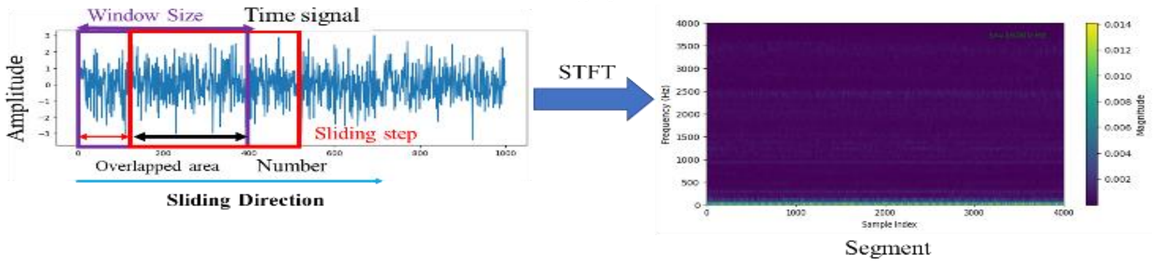

Fig.1. Vibration Time-Frequency Spectrogram Transformation Diagram

Methodology

- The triple-channel VGG19 model comprises 16 convolutional layers and three fully connected layers. It utilizes stacked 3x3 filters to deepen the network for enhanced feature extraction, as illustrated in Fig.2. This study employs transfer learning to analyze vibration and acoustic signals under varying loads for rotating machinery fault diagnosis.

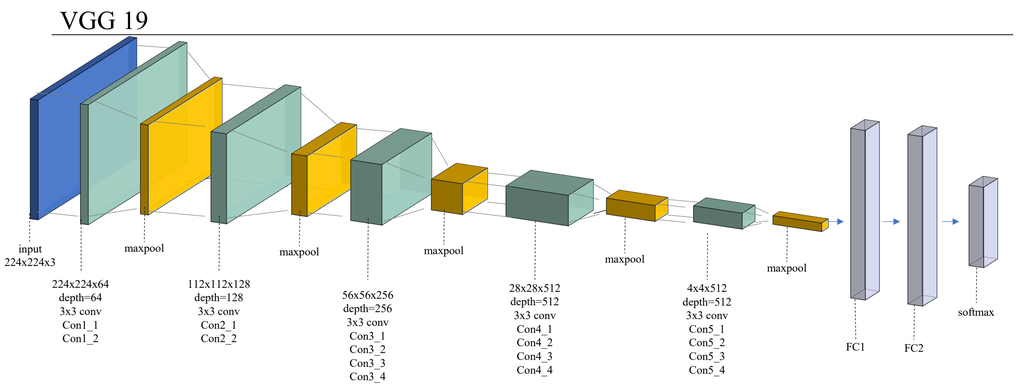

Fig.2. Structure diagram of triple-channel model (VGG19)

- The model is pre-trained on 0 hp data to capture fault features, followed by sequential fine-tuning on 0.5 hp and 1 hp datasets to enhance adaptability. Diagnostic robustness is evaluated using accuracy, precision, and recall.

Sliding sampling and 2D image conversion

- The system samples X-axis vibration and sound signals at 16 kHz, with a single 10 s sampling period (160,000 points per transaction).
- To extract more features, the first step is sliding sampling which involves extracting subsequences from a signal. The window size is set at 4096 points, resulting in each segment having 4096 data points. The sliding step is set at 1024 with an overlap rate of 75% .
- After sliding sampling, the time-frequency domain signals are transformed into 2D images The total sample count $na\ (j)$ for each category is calculated by equation (2).

$$n_a(j) \approx \frac{16000 \times 10 - 4096}{1024} + 1 \ = 153;\ \ j = 0,1,2,\dots 9 \qquad (2)$$

- Subsequently, one time-domain and two time-frequency image are normalized and combined to create a three-channel 64x64 image as shown in Fig.3.

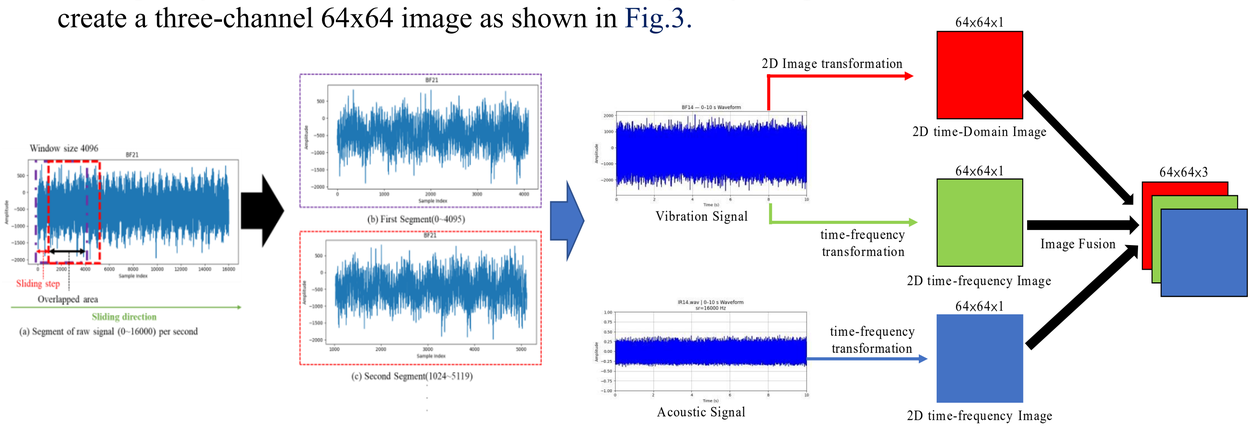

Fig.3: Three-channel image composition diagram

Data Augmentation and Transfer Learning

To improve generalization, vibration signals were augmented with **Gaussian noise**, doubling the dataset without distortion. The model utilizes VGG19 pre-trained on no-load data. The dataset is split into **49%** for training, **21%** for validation, and **30%** for testing. A low learning rate and early stopping were applied to prevent overfitting. During transfer learning, feature layers were frozen, and new fully connected layers were fine-tuned on target datasets.

Experimental Results

To simulate bearing faults under varying loads, a self-built test rig was used at 1750 rpm with three motor loads: no load, 0.5 hp, and 1 hp. The system setup (Fig.4) includes a 1 kW AC servo motor, servo driver, PLC, HMI, magnetic powder brake, interchangeable bearing module (Fig.6), directional microphone, and vibration sensor. The different types of defects are shown in (Fig.5).

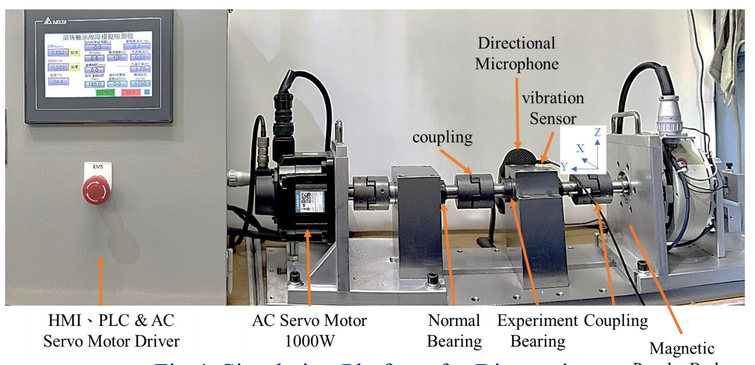

Fig.4. Simulation Platform for Diagnosis

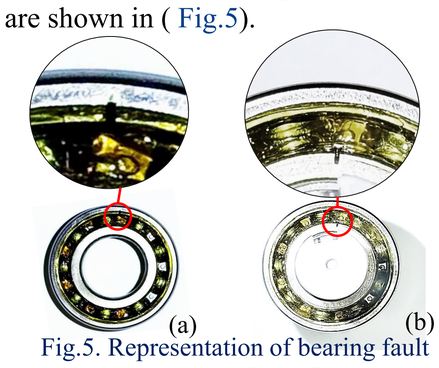

Fig.5. Representation of bearing fault conditions: (a) Outer Race Fault and (b) Inner Race Fault.

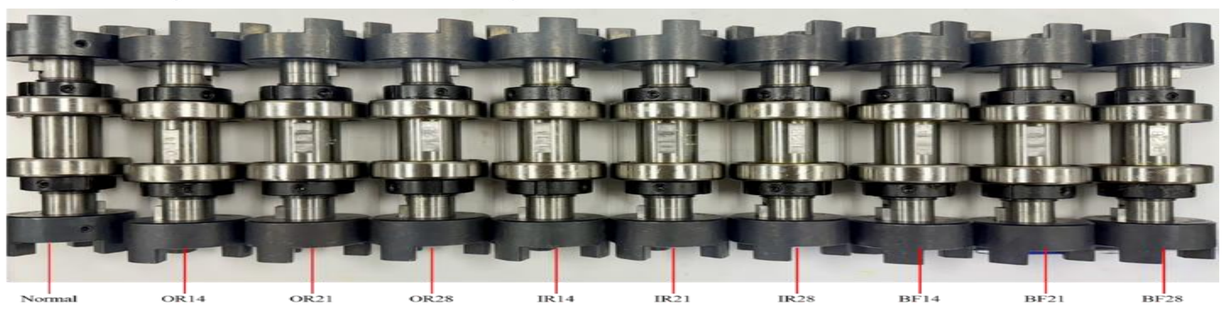

Fig.6. Interchangeable Bearing Modules

This study uses a triple-channel input—vibration time, time-frequency, and acoustic time-frequency images—to capture key fault features and enhance recognition. Transfer learning under 0.5 hp and 1 hp conditions verifies the effectiveness of this multi-source input.

As shown in Fig.7, confusion matrices show accurate classification, while Fig.8 P–R curves confirm model robustness. Training results indicate the pretrained model is stable and suitable for transfer learning.

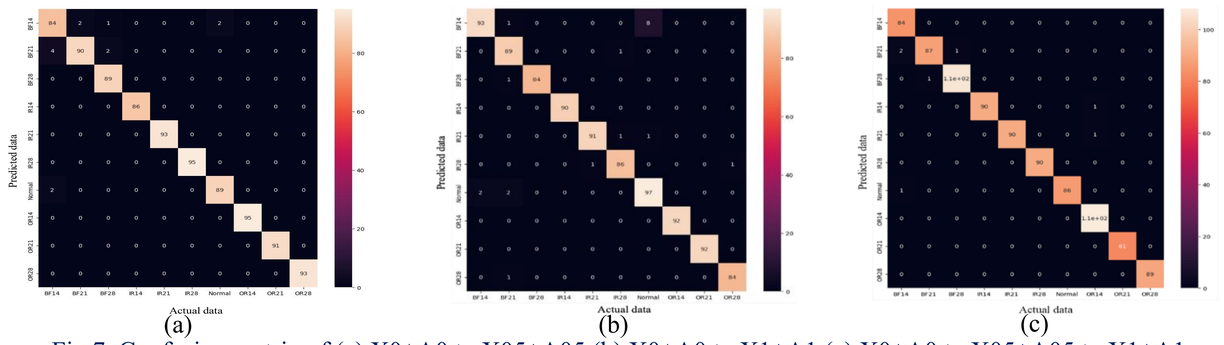

Fig.7. Confusion matrix of (a) X0+A0 to X05+A05 (b) X0+A0 to X1+A1 (c) X0+A0 to X05+A05 to X1+A1

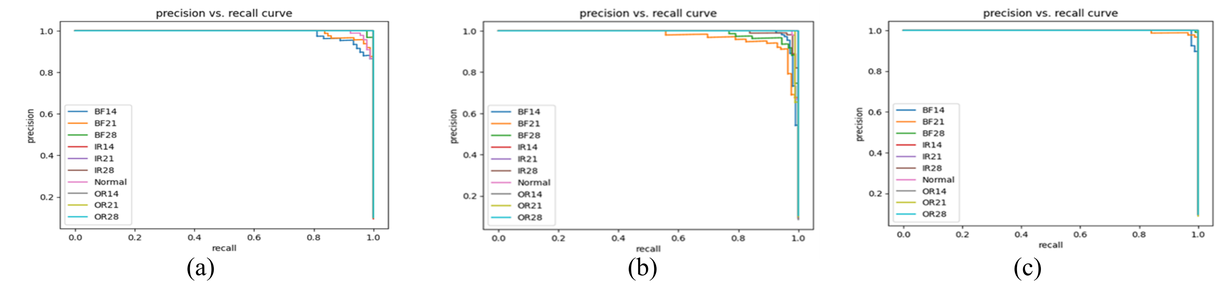

Fig.8 P-R curve of (a) X0+A0 to X05+A05 (b) X0+A0 to X1+A1 (c) X0+A0 to X05+A05 to X1+A1

In transfer learning across different load conditions, all models—whether single-channel, dual-channel, and triple-channel promising performance in both pretraining and transfer phases. Table 1 presents the classification accuracies of models under different input channel configurations. The results indicate that both channel design and load transition strategy significantly affect fault classification accuracy.

Table 1. Accuracy of Transfer Learning with different Channel Inputs

Models with different channel		**Accuracy**	**Training Time(sec）***
Single-Channel (Vib TF) (A TF)	X0 to X05	97.49%	128
	X0 to X1	94.98%	146
	X0 to X05 to X1	97.16%	293
	A0 to A05	89.21%	100
	A0 to A1	87.14%	75
	A0 to A05 to A1	91.06%	83
Dual-Channel (Vib TD+TF) (Vib TF+ A TF)	X0 to X05	96.92%	153
	X0 to X1	97.05%	144
	X0 to X05 to X1	98.07%	229
	X0+A0 to X05+A05	97.93%	104
	X0+A0 to X1+A1	97.05%	191
	X0+A0 to X05+A05 to X1+A1	98.88%	118
Triple-Channel (Vib TD+TF + A TF)	X0+A0 to X05+A05	97.82%	163
	X0+A0 to X1+A1	97.71%	93
	X0+A0 to X05+A05to X1+A1	**99.23%**	153

*The improved training efficiency is attributed to a high-performance 64-bit Windows 11 workstation equipped with an Intel Core i7-13700K CPU, 64 GB of memory, and an NVIDIA RTX 4070 GPU. (Python 3.9.19)

Conclusions

This study proposes a VGG19-based transfer learning framework using 64×64×3 images that fuse vibration and acoustic features. This multi-channel design improves classification by eliminating manual feature extraction. Triple-channel input achieved the highest accuracy of 99.23% in a three-stage transfer (0 → 0.5 → 1 hp), outperforming other configurations. Multi-stage transfer also enhanced adaptability and convergence. Gaussian noise slightly reduced accuracy but improved robustness. Overall, the method enhances cross-domain performance, with future work focusing on broader application and feature refinement.

Acknowledgments

The authors would like to thank National Science and Technology Council, Taiwan for funding this research project (Grant NO: NSTC113-2221-E011-082). Thanks also give to Harbor tech, Taipei, Taiwan for providing vibration sensor (HEA932) and required technical support.