Real-Time Vibration-Based Bearing Fault Diagnosis Under Time-Varying Speed Conditions

4 authors:

Tuomas Jalonen
Tampere University
5 PUBLICATIONS   **7** CITATIONS

SEE PROFILE

Mohammad Al-Sa'd
University of Helsinki
16 PUBLICATIONS   **466** CITATIONS

SEE PROFILE

Serkan Kiranyaz
Qatar University
370 PUBLICATIONS   **16,678** CITATIONS

SEE PROFILE

Moncef Gabbouj
Tampere University
1,029 PUBLICATIONS   **26,739** CITATIONS

SEE PROFILE

Real-Time Vibration-Based Bearing Fault Diagnosis Under Time-Varying Speed Conditions

Tuomas Jalonen*, Mohammad Al-Sa'd*†, Serkan Kiranyaz‡, Moncef Gabbouj*

*Department of Computing Sciences, Tampere University, 33720 Tampere, Finland
Email: tuomas.jalonen@tuni.fi; mohammad.al-sad@tuni.fi; moncef.gabbouj@tuni.fi
†Department of Physiology, University of Helsinki, 00014 Helsinki, Finland
Email: mohammad.al-sad@helsinki.fi
‡Department of Electrical Engineering, Qatar University, 2713 Doha, Qatar
Email: mkiranyaz@qu.edu.qa

Abstract—Detection of rolling-element bearing faults is crucial for implementing proactive maintenance strategies and for minimizing the economic and operational consequences of unexpected failures. However, many existing techniques are developed and tested under strictly controlled conditions, limiting their adaptability to the diverse and dynamic settings encountered in practical applications. This paper presents an efficient real-time convolutional neural network (CNN) for diagnosing multiple bearing faults under various noise levels and time-varying rotational speeds. Additionally, we propose a novel Fisher-based spectral separability analysis (SSA) method to elucidate the effectiveness of the designed CNN model. We conducted experiments on both healthy bearings and bearings afflicted with inner race, outer race, and roller ball faults. The experimental results show the superiority of our model over the current state-of-the-art approach in three folds: it achieves substantial accuracy gains of up to 15.8%, it is robust to noise with high performance across various signal-to-noise ratios, and it runs in real-time with processing durations five times less than acquisition. Additionally, by using the proposed SSA technique, we offer insights into the model's performance and underscore its effectiveness in tackling real-world challenges.

Index Terms—Bearing fault diagnosis, deep learning, varying speed, damage detection, industrial safety.

I. INTRODUCTION

Rolling-element bearings are fundamental components in various industrial machinery and mechanical systems, such as pumps, compressors, and motors, to name a few [1]; see Fig. 1. However, over time, bearings are susceptible to wear, fatigue, and faults that can lead to unexpected machinery failures, production downtime, and costly maintenance [2]. Hence, the precise diagnosis of bearing faults is of utmost importance [2]. Fault diagnosis of rolling-element bearings has been a focal point of interest among researchers over several decades [3]. Data pertinent to this diagnosis can be acquired from various sources, including vibration [3], acoustic emission [4], oil debris [5], electrostatic measurements [6], motor current [7], and visual images [2]. Nevertheless, studies have shown that vibration is one of the most reliable and cost-effective techniques, and allows capturing subtle changes in dynamics that are associated with early-stage defects [2], [7].

The integration of advanced machine learning technologies with vibration analysis has ushered in a new era for bearing
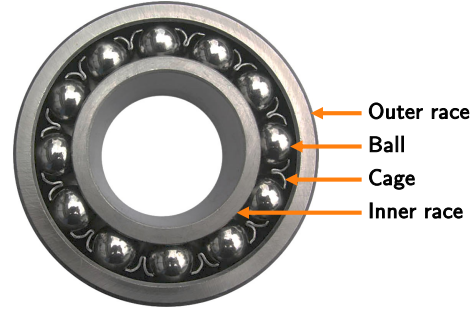

Fig. 1. A rolling element bearing example. The main parts outer race, inner race, balls, and cage are illustrated. Redrawn from https://upload.wikimedia.org/wikipedia/commons/f/f6/HARP_bearing.jpg.

fault diagnosis [1]. Specifically, leveraging Convolutional Neural Networks (CNN) was found to be a promising approach for detecting and identifying bearing defects at their early stages; thereby, enabling proactive maintenance strategies and ultimately reducing the economic and operational impact of unexpected failures [8], [9], [10]. Unfortunately, current solutions are usually designed to work in a very restricted setting and do not integrate the varying conditions and operations found in a real-life scenario, e.g., heavy noise and time-varying rotational speeds. Therefore, we propose a lightweight real-time CNN network to diagnose bearing faults under various noise levels and time-varying rotational speeds. In addition, we design a novel Fisher-based separability analysis technique to rationalize the efficacy of the proposed model. The key contributions of this paper are:

- A simple yet effective CNN architecture that exhibits robustness in noisy and variable speed conditions.
- A precise real-time diagnostic tool for various bearing faults that surpasses the current state-of-the-art technique [11] with a great margin; up to 15.8% gain in accuracy.
- A Fisher-based spectral separability analysis technique that surrogates the model's performance and justifies its class-specific efficacy and shortcomings.

Conventional signal processing techniques involve manual extraction of features for diagnosing bearing defects [2], including time-domain acoustic features [4], spectral amplitude

modulation and kurtosis [12], spectral segmentation [13], variational mode decomposition [14], [15], multi-scale indicators combined with random forest [14] and support vector machines [15]. While hand-crafted features are effective in stable environments, they face challenges in noisy conditions, prompting the adoption of automatic feature extraction techniques, notably CNNs [16]. For instance, a 2D discrete Fourier transform was performed on the vibration signals in [17] to create spectral images, which were used for CNN-based classification. Additionally, deep belief networks [18], [19], physics-informed CNN models [20], [21], and attention mechanisms [22] have all been explored for bearing fault diagnosis. Furthermore, Transformers have demonstrated remarkable fault detection performance but were found computationally intensive and less suitable for smaller datasets [23].

Unfortunately, despite the wide array of methods discussed above, they do not consider the case of time-varying rotational speeds i.e., non-stationary signal analysis [24], and tend to train/test only on fixed working conditions. To remedy this limitation, a Transformer with a self-attention module [25] and recurrent neural networks [26] were suggested to detect rolling-element bearing faults under time-varying speeds. Nevertheless, a new time-varying speed dataset was recently developed at the Korean Advanced Institute of Science and Technology [27] and the current state-of-the-art solution for it is the physics-informed residual network (PIResNet) [11]. Another solution was developed in [28], but the fault diagnosis problem was simplified into a binary classification to detect healthy and non-healthy bearings; hence, it does not qualify as a bearing diagnostic tool. Consequently, comparative evaluations are performed against the state-of-the-art PIResNet approach [11].

II. Methodology

A rolling-element bearing is comprised of rollers that serve to uphold the spatial separation between two bearing races [29]; see Fig. 1. The primary function of a ball bearing is to mitigate rotational friction while providing support for both radial and axial loads [29]. This is accomplished by employing a minimum of two races to encase the balls and facilitate the transmission of loads [29]. Faults in bearings represent the predominant category of anomalies occurring in induction machines; hence, continuous monitoring is essential [30]. Fortunately, vibration in machines is mainly generated due to defective bearings; hence, analyzing its morphology allows us to diagnose the bearing's health state [17], [31].

A. Dataset description

We utilized an open-access dataset collected from a bearing test rig at the Korean Advanced Institute of Science and Technology (KAIST) [27]. The test rig consists of a three-phase induction motor, torque meter, gearbox, bearing housing A, bearing housing B, rotors, and a hysteresis brake [27]. Four ceramic shear ICP-based accelerometers were mounted on the x- and y-directions of the two bearing housings [27]. The motor was operated with time-varying speeds (680-2460

rpm) and the bearing in housing B was seeded with outer race, inner race, and bearing ball faults. These experiments yielded four classes named *Normal*, *Outer*, *Inner*, and *Ball*, describing the bearing's normal operation and three common faults. The vibration signals under time-varying speed conditions were acquired by the accelerometers installed on housing B. The sampling frequency was set to 25.6 kHz and 2,100 seconds of vibration data were collected for each of the four classes.

B. Signal preprocessing

The vibration signals from the two sensors were first downsampled to 20 kHz using a finite impulse response anti-aliasing lowpass filter with delay compensation ($f_s = 20000$). This operation yields a discrete multi-sensor signal $\boldsymbol{s} = [s_1(n), s_2(n)]^T$ of size $2 \times N$ where $s_q(n) = [s_q(1), s_q(2), \cdots, s_q(N)]$ is the signal acquired by sensor q with $N = 42 \times 10^6$ time samples [32]. After that, we formed a noisy multi-sensor vibration signal $\boldsymbol{x}$ as:

$$\boldsymbol{x} = \boldsymbol{s} + \alpha \, \boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{\eta} = [\eta_1(n), \eta_2(n)]^T$ holds independent and identically distributed real-valued noise, $\eta_q(n) \sim \mathcal{N}(0, 1)$ is a zero-mean white Gaussian noise with unit variance, and $\alpha = [\alpha_1, \alpha_2]^T$ are factors that control the amount of noise to be added to $\boldsymbol{s}$ given a pre-defined signal-to-noise ratio (SNR), i.e.:

$$\alpha_q = \sqrt{\frac{10^{-\text{SNR}/10}}{N} \sum_{n=1}^{N} s_q^2(n)}. \tag{2}$$

In this study, we conducted experiments utilizing both clean vibration signals and those subjected to various degrees of noise-induced degradation. Specifically, we corrupted the vibration waveforms with noise at the following SNR levels: 20 dB, 15 dB, 10 dB, 5 dB, 0 dB, and -5 dB, to span the range of mild to moderate to severe noise conditions.
Furthermore, we divided the temporal course of the signals in $\boldsymbol{x}$ into smaller segments of length L with no overlap to prevent any inadvertent data leakage when training and/or testing the CNN model. The segmented signal $\boldsymbol{v}$ is expressed as a 3-way tensor of size $2 \times P \times L$, i.e.:

$$\boldsymbol{v} = \begin{bmatrix} v_{1,1}(\ell) & v_{1,2}(\ell) & \cdots & v_{1,P}(\ell) \\ v_{2,1}(\ell) & v_{2,2}(\ell) & \cdots & v_{2,P}(\ell) \end{bmatrix}, \tag{3}$$

$$v_{q,p}(\ell) = [x_q\left(L(p-1)+1\right), x_q\left(L(p-1)+2\right), \cdots, x_q\left(pL\right)], \tag{4}$$

where $P = N/L$ is the total number of segments with length L, $p \in [1, P]$, and $\ell \in [1, L]$. The segment length L is a paramount parameter that needs careful selection. On the one hand, it needs to be small enough to generate a sufficient number of samples for training and to minimize additional complexities emerging from the motor's varying speed. On the other hand, it should be wide enough to include the motor's vibration complexities and to learn class-specific patterns. To determine an adequate segment length, we conducted a frequency analysis on the motor's speed to identify its maximum rate of change; see Fig. 2. The analysis showed that the motor's rotational speed varies at two specific frequencies; 8 Hz and
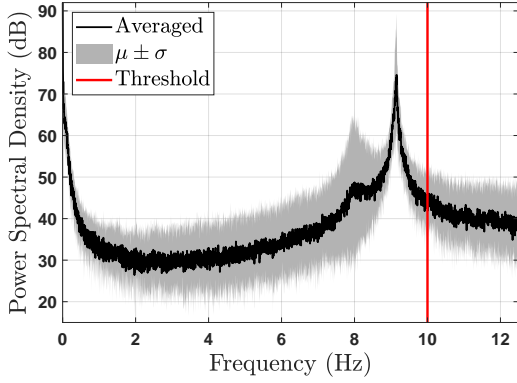
Fig. 2. Frequency analysis of the motor's rotational speed. The power spectral density was estimated by the nonuniform Fourier transform (NUFT) because of the speed's inconsistent acquisition rate. Frequency scaling for the NUFT was performed using a sampling frequency of 12.5 Hz; the reciprocal of the shortest acquisition time. The results were averaged across the different classes and dataset files and plotted with their 68.3% confidence intervals.

9.15 Hz. Therefore, we decided 10 Hz, or 0.1 seconds, is a suitable maximum rate of range which includes 99.6% of the total power spectrum; almost all rapid changes in rotational speed. The selected threshold corresponds to a segment length of 2,000 samples ($L = 2000$) which yields a total of 21,000 segments ($P = 21000$) which is ample for training and/or testing deep learning models.

Finally, the generated segments in v exhibit variations in scale. Therefore, standardization is imperative for rigorous and precise processing. In this study, we calculated the standardized, or z-score normalized, tensor z as follows:

$$z = \begin{bmatrix} z_{1,1}(\ell) & z_{1,2}(\ell) & \cdots & z_{1,P}(\ell) \\ z_{2,1}(\ell) & z_{2,2}(\ell) & \cdots & z_{2,P}(\ell) \end{bmatrix}, \quad (5)$$

$$z_{q,p}(\ell) = \left(v_{q,p}(\ell) - \mu_p \right) / \sigma_p, \quad (6)$$

where μ_p and σ_p are the mean and standard deviation of the p-th segment from both sensors (flattened), respectively.

C. The deep learning model

We designed a lightweight CNN model to allow the utility of inexpensive edge computing units for deployment. The network accepts the standardized vibration signals z coming from both accelerometers (mounted on the x- and y-directions) of housing B, i.e., the network's input size is 2000×2 (input signal is transposed). Afterward, the input signals are passed through five convolutional blocks each containing a convolutional layer for feature extraction with a ReLU activation function to account for non-linear effects, and a max-pooling layer of size 2×1 to summarize the feature maps. The convolution layer in the first block has a kernel size of 5×1 while the rest have 3×1 kernels. The extracted features (coming from the last convolutional block) are flattened and passed through a dense layer with 128 nodes and a ReLU activation function. Finally, they are fed to a dropout layer with a rate of 0.4 and a classification layer with 4 nodes and a Softmax activation function to transform the network's raw

outputs into a vector of probabilities. In total, the model has 558,660 trainable parameters.

Training and testing the CNN model was conducted in a 5-fold cross-validation to accurately estimate its ability to generalize to an independent dataset. First, we divided the 84,000 samples (four classes each having 21,000 segments) into five equally sized shuffled splits and maintained their class balance (stratified folds). Out of the five splits, a single split was kept for testing, while the remaining four splits were used for training. We repeated this process five times, with each of the five splits used exactly once for testing. Specifically, each model was trained with 67,200 samples and tested with 16,800 samples. Training underwent for 150 epochs, to ensure convergence, and Adam optimizer was employed to minimize the categorical cross-entropy loss [33], i.e.:

$$\mathcal{L} = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}), \quad (7)$$

where $\mathcal{L}$ denotes the loss, y and $\hat{y}$ are the true and predicted class labels, respectively. We set the batch size to 100, and the learning rate to 10^{-5} for epochs 1-101 and 10^{-6} for epochs 102-150. Finally, we tested the CNN model using the weights from the last epoch (no early stopping), and the same structure and training settings were used for all noise levels.

D. The spectral separability analysis

The rationale for the efficacy of a deep learning model or elucidation of the underlying factors contributing to its shortcomings is imperative for its seamless integration and advancement. In this study, we propose a Fisher-based spectral separability analysis (SSA) to quantify the presence of discriminative features in the vibration signals. The SSA facilitates differentiating the distinct bearing health states and serves as a surrogate indicator for the model's performance. For instance, if there is a spectral distinction between two classes, the model should be capable of detecting it, whereas if such disparity is absent, the model's expected/unexpected limitations/efficacy becomes apparent. The SSA starts by calculating the smoothed power spectral density (PSD) of the vibration signal $z_{q,p}^{(i)}(\ell)$ belonging to class i by:

$$\hat{z}_{q,p}^{(i)}(k) = 10 \log_{10} \left(\frac{1}{L-1} \left| \mathcal{F}_{\ell \to k} \{ z_{q,p}^{(i)}(\ell) \} \right|^2 \underset{k}{*} \omega_h(k) \right), \quad (8)$$

where $\hat{z}_{q,p}^{(i)}(k)$ is the PSD of $z_{q,p}^{(i)}(\ell)$, $\mathcal{F}_{\ell \to k}$ is the Fourier transform operator from the discrete time ℓ to discrete frequency k, $\underset{k}{*}$ denotes linear convolution in k, and $\omega_h(k)$ is a smoothing window of length h such that $\omega_h(k) \neq 0$ for $|k| \leq h/2$, $\omega_h(k) = \omega_h(-k)$, and $\omega_h(k_1) > \omega_h(k_2)$ for $|k_1| < |k_2|$. After that, the Fisher criterion (FC) for measuring the separability between classes' i and j power spectra is expressed as:

$$f_q^{(i,j)}(k) = \frac{|\mu_{q,i}(k) - \mu_{q,j}(k)|^2}{\sigma_{q,i}^2(k) + \sigma_{q,j}^2(k)}, \quad (9)$$

where $\mu_{q,i}(k)$ and $\sigma_{q,i}^2(k)$ are the mean and variance of $\hat{\boldsymbol{z}}_{q,p}^{(i)}(k)$, respectively, i.e.:

$$\mu_{q,i}(k) = \frac{1}{P}\sum_{p=1}^{P}\hat{z}_{q,p}^{(i)}(k). \tag{10}$$

$$\sigma_{q,i}^2(k) = \frac{1}{P}\sum_{p=1}^{P}\left(\hat{z}_{q,p}^{(i)}(k) - \mu_{q,i}(k)\right)^2. \tag{11}$$

Finally, the FCs are thresholded and smoothed to retain highly separable continuous frequency bands, i.e.:

$$\zeta_q^{(i,j)}(k) = \left(\begin{cases} f_q^{(i,j)}(k) & : f_q^{(i,j)}(k) > \epsilon \\ 0 & : \text{otherwise} \end{cases}\right) \underset{k}{*}\, \omega_h(k), \tag{12}$$

where we set the threshold ϵ to 2 to allow a maximum of 31.73% overlap between the classes' probability distributions.

E. Performance evaluation and analysis

The performance of the CNN model was evaluated by employing a one-versus-all approach, thereby transforming the initial multi-class problem into a set of binary classification tasks. The assessment of each binary task was quantified by accuracy, precision, recall, and F1-score, and then these metrics were averaged across all binary tasks (macro-averaged) to assess the multi-class performance, i.e.:

$$\text{Accuracy} = \frac{1}{I}\sum_{i=1}^{I}\left(\frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}\right), \tag{13}$$

$$\text{Precision} = \frac{1}{I}\sum_{i=1}^{I}\left(\frac{TP_i}{TP_i + FP_i}\right), \tag{14}$$

$$\text{Recall} = \frac{1}{I}\sum_{i=1}^{I}\left(\frac{TP_i}{TP_i + FN_i}\right), \tag{15}$$

$$\text{F1-Score} = \frac{1}{I}\sum_{i=1}^{I}\left(\frac{2\,TP_i}{2\,TP_i + FN_i + FP_i}\right), \tag{16}$$

where $I = 4$ is the total number of classes, TP_i, TN_i, FP_i, and FN_i are true positives, true negatives, false positives, and false negatives, respectively, when detecting class i and rejecting all other classes.

Moreover, we analyzed the predictive capacity of the CNN model by t-distributed stochastic neighbor embedding (t-SNE); a dimensionality reduction method that clusters similar high dimensional samples and departs dissimilar ones in 2- or 3-dimensional space [34]. This technique helps in characterizing the model's predictive power when supplied with new vibration signals. In other words, given large enough training samples, if the t-SNE shows clear inter-class separability, one may infer the adequacy of the model for unseen samples.

Finally, the complexity of the CNN model was assessed by its inference time. We conducted Monte-Carlo simulations where we fed the model with 1000 test samples, predicted their class, and repeated the process 10 times for validation. We used a MacBook Pro with ARM-based M1 Pro chip, integrated 16-core GPU, 16-core neural engine, and 16 GB of RAM.
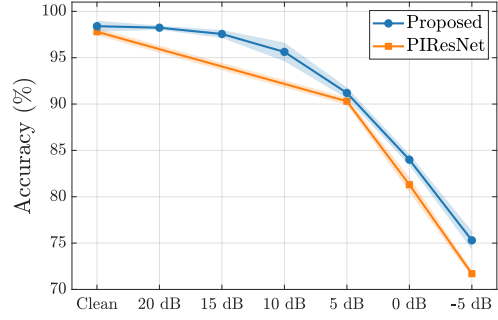

Fig. 3. Comparing the testing macro-averaged accuracy of the CNN network to the current state-of-the-art technique, the PIResNet [11], using both clean and noisy vibration signals. The accuracy curves are plotted at varying SNR levels along with their 95% confidence intervals ($\mu \pm 2\sigma$).

III. RESULTS AND DISCUSSION

A. Fault diagnosis performance

Fig. 3 demonstrates the testing accuracy of the designed CNN network for classifying the four health states of the bearing across the different SNR levels. Additionally, it illustrates a comparison between our proposed model and the current state-of-the-art approach, PIResNet [11]. The results indicate that, whether in clean or severe noisy signal conditions, the proposed approach surpasses the performance of the PIResNet (up to a 3.6% point gain in average accuracy). The extent of this performance enhancement inversely correlates with the SNR level, meaning that the superiority of the proposed approach becomes more pronounced as the level of noise in the vibration signal increases. Table I confirms this observation by extending the comparison to include the complete set of testing metrics for both models; accuracy, precision, recall, and the F1-score. Specifically, it reveals a persistent outperformance of the proposed CNN in comparison to the PIResNet across all metrics and SNR levels, thus rendering the additional complexity associated with the PIResNet redundant.

Furthermore, Fig. 4 details the testing performance by presenting the confusion matrices when using clean and noisy vibration signals at three SNR levels: 5 dB, 0 dB, and -5 dB. The results reveal that: (1) the *Inner* class is the easiest to predict as it had the least amount of accuracy drop; from 100% for clean signals to 95.6% for noisy signals at -5 dB SNR; (2) differentiating the *Normal* and *Outer* states becomes particularly challenging at low SNR, due to the morphological features of these states being predominantly indistinguishable noise; (3) a moderate misclassification between the *Normal* and *Ball* states exists at -5 dB; and (4) by examining the confusion matrices' progression through the SNR levels, one notes a skewness towards the *Inner* and *Ball* states that increases with the amount of noise. This performance imbalance suggests that the temporal morphology of these states is mainly impulsive; hence, they are less affected by additive white Gaussian noise. Finally, when contrasting the results in Fig. 4 with those reported in [11], it becomes apparent that the designed

TABLE I
THE TESTING PERFORMANCE IN TERMS OF MACRO-AVERAGED ACCURACY, PRECISION, RECALL, AND F1-SCORE. THE METRICS ARE SUMMARIZED BY THEIR 5-FOLD AVERAGED PERCENTAGES ± STANDARD DEVIATIONS AND THE BEST OUTCOMES ARE HIGHLIGHTED IN BOLD. THE PRECISION, RECALL, AND F1-SCORE OF PIRESNET ARE CALCULATED FROM THE REPORTED BEST-PERFORMING CONFUSION MATRICES IN [11]; THUS, THE RESULTS' STANDARD DEVIATIONS ARE MISSING.

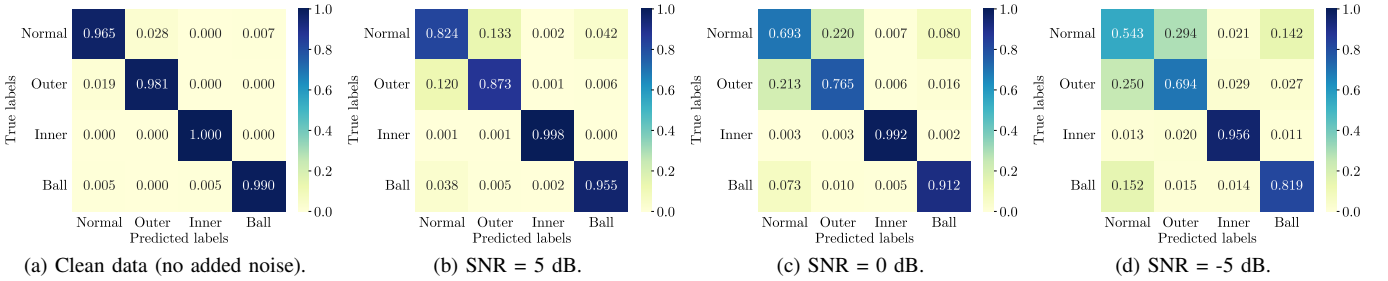|	Clean		SNR = 5 dB		SNR = 0 dB		SNR = -5 dB	
	Proposed	PIResNet [11]	Proposed	PIResNet [11]	Proposed	PIResNet [11]	Proposed	PIResNet [11]
Accuracy	98.4 ± 0.3	97.8 ± 0.2	91.2 ± 0.3	90.3 ± 0.2	84.0 ± 0.3	81.3 ± 0.4	75.3 ± 0.5	71.7 ± 0.2
Precision	98.4 ± 0.3	$98.0 \pm$ N/A	91.3 ± 0.4	$90.5 \pm$ N/A	84.0 ± 0.2	$81.6 \pm$ N/A	75.1 ± 0.7	$72.3 \pm$ N/A
Recall	98.4 ± 0.3	$98.0 \pm$ N/A	91.2 ± 0.4	$90.5 \pm$ N/A	84.0 ± 0.3	$81.6 \pm$ N/A	75.3 ± 0.6	$72.0 \pm$ N/A
F1-Score	98.4 ± 0.3	$98.0 \pm$ N/A	91.2 ± 0.4	$90.5 \pm$ N/A	84.0 ± 0.3	$81.6 \pm$ N/A	75.2 ± 0.7	$72.2 \pm$ N/A

Fig. 4. The testing confusion matrices averaged over the five data splits. The results depict the performance when using (a) clean, and (b)-(d) noisy vibration signals at the following SNR levels: (b) 5 dB, (c) 0 dB, and (d) -5 dB.

CNN outruns the PIResNet with a significant performance gain of up to 15.8% points. Nevertheless, our results are averaged over the five cross-validation splits while PIResNet results are drawn from the best-performing confusion matrices. Therefore, the former comparison constitutes a lower limit on the performance gain we can achieve.

B. The t-SNE and complexity analysis

Fig. 5 shows the t-SNE results which provide insight into the learned features of the CNN model, where they show the model's ability to learn highly discriminative features when supplied with clean data, and its resilience to noise when distinguishing some of the bearing's health states. On the one hand, the results in Fig. 5a reveal a clear inter-class separation among all health states when using clean vibration signals. On the other hand, Figs. 5b-d demonstrate that inter-class separability diminishes in correlation to the level of introduced noise. In particular, we found that the *Inner* class features form the most isolated cluster at all SNR levels, the feature overlap between the *Normal* and *Ball* classes increases with noise, and that features from the states *Normal* and *Outer* intertwine to become indistinguishable from each other when noise is introduced. The latter behavior is expected as it is a consequence of the noted model's performance imbalance at low SNR levels; see Fig. 4d. Moreover, the model inference time statistics are demonstrated in Fig. 6. The results indicate an average inference time of 20.2 ms with a 0.36 standard deviation. This corresponds to a processing time approximately five times less than the duration of data

acquisition. In other words, it takes $\approx$ 20 ms to process 100 ms of vibration data coming from the two sensors.

C. The spectral separability analysis

Fig. 7 illustrates the outcomes of the SSA conducted on both the uncontaminated and noisy (SNR = 0 dB) vibration signals, offering a method for explaining the CNN model's efficacy and pitfalls. In the context of clean data, as displayed in Fig. 7a, the findings reveal the following marked differences in spectral patterns: (1) the spectral characteristics of the *Inner* class exhibit pronounced dissimilarity when compared to all other classes, notably within the frequency bands $0 - 0.3$ kHz, $3.8 - 5$ kHz, $5.8 - 6.2$ kHz, $7.3 - 7.5$ kHz, $8.9 - 9.2$ kHz, and $9.6 - 9.8$ kHz; (2) the *Ball* class displays moderate spectral distinctions relative to the *Normal* and *Outer* classes, as evident around $6.4 - 6.8$ kHz; and (3) the *Normal* and *Outer* classes exhibit slight difference around $9.4 - 10$ kHz. This implies that the high performance demonstrated in Fig. 4a is a consequence of the model's efficacy in acquiring the temporal and spectral knowledge of the non-stationary vibration signals. On the other hand, in the case of noisy data, as shown in Fig. 7b, the SSA results show that: (1) no significant inter-class separability exists apart from one group, the *Inner* class, as evident within $5.8 - 6.2$ kHz and $8.9 - 9.2$ kHz; and (2) the separability of the *Inner* class became less significant when compared to the clean case. These observations explain the model's shortcomings at low SNR levels and its performance imbalance in three folds. First, the *Normal* and *Outer* classes had little discriminative information in the frequency domain; hence, the model relied mainly on their temporal morphology.

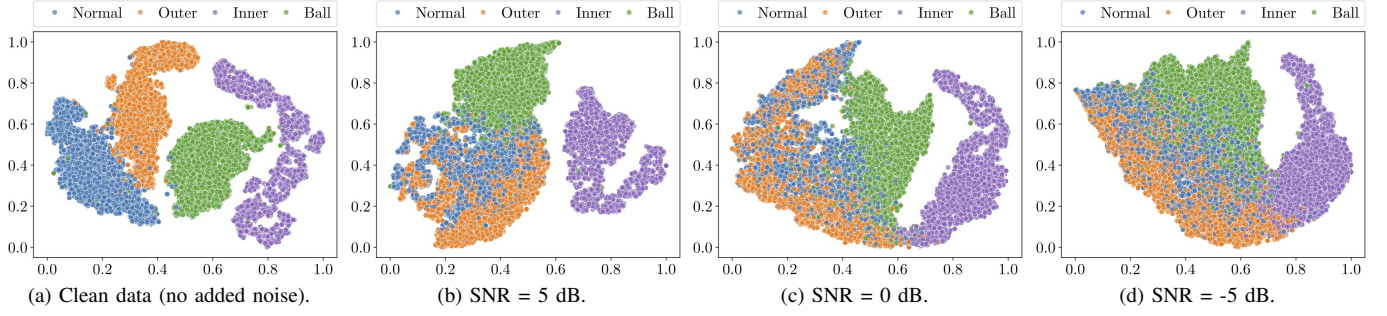(a) Clean data (no added noise). (b) SNR = 5 dB. (c) SNR = 0 dB. (d) SNR = -5 dB.

Fig. 5. The model's testing t-SNE from the first data split. The results depict the t-SNE when using (a) clean, and (b)-(d) noisy data at the following SNR levels: (b) 5 dB, (c) 0 dB, and (d) -5 dB. We used the model's last layer features as input to the t-SNE algorithm.
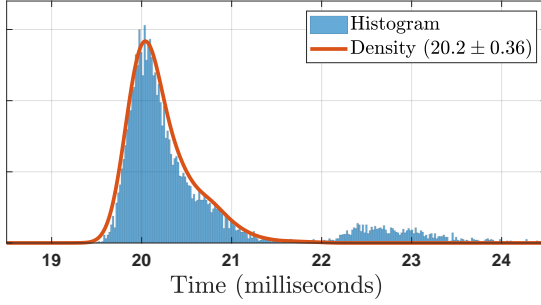

Fig. 6. Statistics of the proposed model's inference time.

However, this dependence becomes unreliable when noise is introduced; therefore, features for these two classes become intertwined and dominate the errors as shown in Fig. 4c and Fig. 5c. Moreover, the *Inner* state had ample spectral disparity or unique features, which allowed high performance even at low SNR levels. Finally, the *Ball* class had moderate discriminative features in the frequency domain; therefore, the model had to rely on both its temporal and spectral morphology, resulting in a reasonable performance drop at low SNR. Nevertheless, this suggests that vibration signals should be represented in the joint time-frequency domain to exploit their degrees of freedom fully and to design highly robust bearing fault detectors.

IV. CONCLUSIONS

Diagnosing faults in bearings at their early stages allows proactive maintenance strategies and reduces the economic and operational impact of unexpected failures. Unfortunately, current fault diagnosis techniques are often tested in restricted settings and do not integrate the variable conditions and operations found in a real-life scenario. In this paper, we proposed a vibration-based efficient real-time convolutional neural network (CNN) to diagnose different bearing faults under various noise levels and time-varying rotational speeds. Besides, we designed a novel Fisher-based spectral separability analysis (SSA) technique to rationalize the efficacy of the proposed model. The experimental results showed that the proposed method outperforms PIResNet, the current state-of-the-art technique, across multiple signal-to-noise ratios with
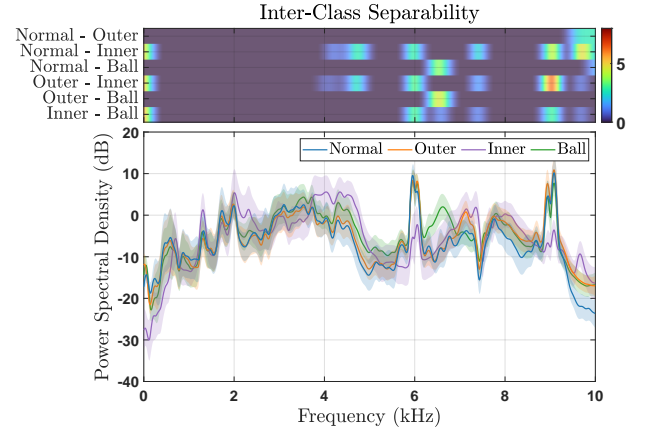
(a) Clean data (no added noise).
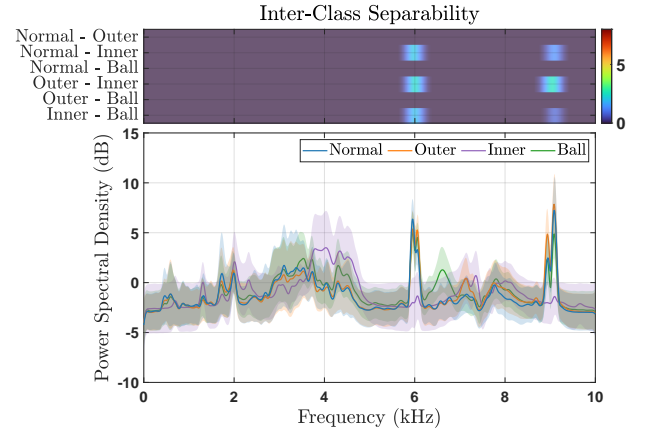
(b) SNR = 0 dB.

Fig. 7. The SSA of the y-directional vibrations (a) without noise, and (b) with noise at 0 dB SNR. The Fourier transforms were calculated with 1024 samples and smoothed via a Gaussian-weighted moving average filter of length 16 samples. The averaged PSDs are shown for each class along with their 68.3% confidence intervals ($\mu \pm \sigma$). Inter-class separability was estimated for each pair of groups in a one-versus-one fashion. The SSA estimates were smoothed by a Gaussian-weighted moving average filter of length 64 samples to highlight wide rather than narrow separable bands.

accuracy gains reaching up to 15.8% while running in real-time. Moreover, through the SSA outcomes, we offered an explanation for the model's behavior under heavy noise and

indicated the need for representing the vibration signals in the joint time-frequency domain to design robust bearing diagnostic tools. Our future research directions include testing the proposed CNN on different datasets, designing a time-frequency-based bearing fault diagnosis system, and incorporating advanced techniques such as operational neural networks [35] with the recent neuron models [36] to improve the model's learning capacity.

REFERENCES

[1] O. AlShorman, M. Irfan, N. Saad, D. Zhen, N. Haider, A. Glowacz, and A. AlShorman, "A Review of Artificial Intelligence Methods for Condition Monitoring and Fault Diagnosis of Rolling Element Bearings for Induction Motor," *Shock and Vibration*, vol. 2020, p. 8843759, 2020.

[2] A. Rai and S. H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribology International*, vol. 96, pp. 289–306, 2016.

[3] R. S. Gunerkar, A. K. Jalan, and S. U. Belgamwar, "Fault diagnosis of rolling element bearing based on artificial neural network," *Journal of Mechanical Science and Technology*, vol. 33, no. 2, pp. 505–511, 2019.

[4] A. Aasi, R. Tabatabaei, E. Aasi, and S. M. Jafari, "Experimental investigation on time-domain features in the diagnosis of rolling element bearings by acoustic emission," *Journal of Vibration and Control*, vol. 28, no. 19-20, pp. 2585–2595, 2022.

[5] Y. Zhao, X. Wang, S. Han, J. Lin, and Q. Han, "Fault Diagnosis for Abnormal Wear of Rolling Element Bearing Fusing Oil Debris Monitoring," *Sensors*, vol. 23, no. 7, 2023.

[6] L. Li, H. Hu, and K. Tang, "A specially-designed electrostatic sensor for the condition monitoring of rolling bearings," *Measurement Science and Technology*, vol. 32, no. 3, p. 035110, 2020.

[7] T. Ince, J. Malik, O. C. Devecioglu, S. Kiranyaz, O. Avci, L. Eren, and M. Gabbouj, "Early Bearing Fault Diagnosis of Rotating Machinery by 1D Self-Organized Operational Neural Networks," *IEEE Access*, vol. 9, pp. 139 260–139 270, 2021.

[8] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7067–7075, 2016.

[9] L. Eren, T. Ince, and S. Kiranyaz, "A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier," *Journal of Signal Processing Systems*, vol. 91, no. 2, pp. 179–189, 2019.

[10] O. Abdeljaber, S. Sassi, O. Avci, S. Kiranyaz, A. A. Ibrahim, and M. Gabbouj, "Fault Detection and Severity Identification of Ball Bearings by Online Condition Monitoring," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 10, pp. 8136–8147, 2019.

[11] Q. Ni, J. C. Ji, B. Halkon, K. Feng, and A. K. Nandi, "Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics," *Mechanical Systems and Signal Processing*, vol. 200, p. 110544, 2023.

[12] A. Moshrefzadeh, "Condition monitoring and intelligent diagnosis of rolling element bearings under constant/variable load and speed conditions," *Mechanical Systems and Signal Processing*, vol. 149, p. 107153, 2021.

[13] K. Zhang, Y. Xu, Z. Liao, L. Song, and P. Chen, "A novel Fast Entrogram and its applications in rolling bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 154, p. 107582, 2021.

[14] Q. Hu, X.-S. Si, Q.-H. Zhang, and A.-S. Qin, "A rotating machinery fault diagnosis method based on multi-scale dimensionless indicators and random forests," *Mechanical Systems and Signal Processing*, vol. 139, p. 106609, 2020.

[15] J. Zhou, M. Xiao, Y. Niu, and G. Ji, "Rolling bearing fault diagnosis based on WGWOA-VMD-SVM," *Sensors*, vol. 22, no. 16, p. 6281, 2022.

[16] T. Jalonen, M. Al-Sa'd, R. Mellanen, S. Kiranyaz, and M. Gabbouj, "Real-Time Damage Detection in Fiber Lifting Ropes Using Convolutional Neural Networks," *arXiv*, 2023.

[17] A. Youcef Khodja, N. Guersi, M. N. Saadi, and N. Boutasseta, "Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks," *The International Journal of Advanced Manufacturing Technology*, vol. 106, pp. 1737–1751, 2020.

[18] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Measurement Science and Technology*, vol. 26, no. 11, p. 115002, 2015.

[19] G. Niu, X. Wang, M. Golda, S. Mastro, and B. Zhang, "An optimized adaptive PReLU-DBN for rolling element bearing fault diagnosis," *Neurocomputing*, vol. 445, pp. 26–34, 2021.

[20] M. Sadoughi and C. Hu, "Physics-Based Convolutional Neural Network for Fault Diagnosis of Rolling Element Bearings," *IEEE Sensors Journal*, vol. 19, no. 11, pp. 4181–4192, 2019.

[21] S. Shen, H. Lu, M. Sadoughi, C. Hu, V. Nemani, A. Thelen, K. Webster, M. Darr, J. Sidon, and S. Kenny, "A physics-informed deep learning approach for bearing fault detection," *Engineering Applications of Artificial Intelligence*, vol. 103, p. 104295, 2021.

[22] S. Plakias and Y. S. Boutalis, "Fault detection and identification of rolling element bearings with Attentive Dense CNN," *Neurocomputing*, vol. 405, pp. 208–217, 2020.

[23] Y. Jin, L. Hou, and Y. Chen, "A Time Series Transformer based method for the rotating machinery fault diagnosis," *Neurocomputing*, vol. 494, pp. 379–395, 2022.

[24] M. Al-Sa'd, B. Boashash, and M. Gabbouj, "Design of an Optimal Piece-Wise Spline Wigner-Ville Distribution for TFD Performance Evaluation and Comparison," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3963–3976, 2021.

[25] F. Chi, X. Yang, S. Shao, and Q. Zhang, "Bearing Fault Diagnosis for Time-Varying System Using Vibration–Speed Fusion Network Based on Self-Attention and Sparse Feature Extraction," *Machines*, vol. 10, no. 10, p. 948, 2022.

[26] Z. An, S. Li, J. Wang, and X. Jiang, "A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network," *ISA transactions*, vol. 100, pp. 155–170, 2020.

[27] 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 in brief*, vol. 48, pp. 109 049–109 049, 2023.

[28] S. Anbalagan, D. Agarwal, B. Natarajan, and B. Srinivasan, "Foundational Models for Fault Diagnosis of Electrical Motors," *arXiv*, 2023.

[29] A. Kumar and R. Kumar, "Role of Signal Processing, Modeling and Decision Making in the Diagnosis of Rolling Element Bearing Defect: A Review," *Journal of Nondestructive Evaluation*, vol. 38, no. 1, p. 5, 2018.

[30] K. C. Deekshit Kompella, M. Venu Gopala Rao, and R. Srinivasa Rao, "Bearing fault detection in a 3 phase induction motor using stator current frequency spectral subtraction with various wavelet decomposition techniques," *Ain Shams Engineering Journal*, vol. 9, no. 4, pp. 2427–2439, 2018.

[31] C. Malla and I. Panigrahi, "Review of Condition Monitoring of Rolling Element Bearing Using Vibration Analysis and Other Techniques," *Journal of Vibration Engineering & Technologies*, vol. 7, no. 4, pp. 407–414, 2019.

[32] B. Boashash, A. A. El-Bey, and M. F. Al-Sa'd, "Multisensor Time–Frequency Signal Processing MATLAB package: An analysis tool for multichannel non-stationary data," *SoftwareX*, vol. 8, pp. 53–58, 2018, digital Signal Processing & SoftwareX - Joint Special Issue on Reproducible Research in Signal Processing.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[34] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.

[35] S. Kiranyaz, T. Ince, A. Iosifidis, and M. Gabbouj, "Operational neural networks," *Neural Computing and Applications*, vol. 32, no. 11, pp. 6645–6668, 2020.

[36] S. Kiranyaz, J. Malik, H. B. Abdallah, T. Ince, A. Iosifidis, and M. Gabbouj, "Self-organized Operational Neural Networks with Generative Neurons," *Neural Networks*, vol. 140, pp. 294–308, 2021.