IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 September 2023, accepted 12 October 2023, date of publication 16 October 2023, date of current version 25 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3324961

RESEARCH ARTICLE

Multi-Channel Time-Frequency Domain Deep CNN Approach for Machinery Fault Recognition Using Multi-Sensor Time-Series

RAKESH REDDY YAKKATI[1], SREENIVASA REDDY YEDURI[2], (Member, IEEE), RAJESH KUMAR TRIPATHY[3], (Member, IEEE), AND LINGA REDDY CENKERAMADDI[2], (Senior Member, IEEE)

[1]Department of Mathematics, Birla Institute of Technology, Mesra, Ranchi 835215, India
[2]ACPS Research Group, Department of Information and Communication Technology, University of Agder, 4879 Grimstad, Norway
[3]Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, Hyderabad 333031, India

Corresponding author: Linga Reddy Cenkeramaddi (linga.cenkeramaddi@uia.no)

This work was supported by the Indo-Norwegian Collaboration in Autonomous Cyber-Physical Systems (INCAPS) of the International Partnerships for Excellent Education and Research (INTPART) Program from the Research Council of Norway under Project 287918.

ABSTRACT In the industry, machinery failure causes catastrophic accidents and destructive damage to the machines. It causes the machinery to stop and reduces production, causing financial losses to the industry. As a result, identifying machine faults at an early stage is critical. With the rapid advancement in artificial intelligence-based methods, developing automated systems that can diagnose machinery faults is necessary and challenging. This paper proposes a multi-channel time-frequency domain deep convolutional neural network (CNN)-based approach for machinery fault diagnosis using multivariate time-series data from multisensors (tachometer, microphone, underhang bearing accelerometer, and overhand bearing accelerometer). The wavelet synchro-squeezed transform (WSST) based technique is used to evaluate the time-frequency images from the multivariate time-series data. The time-frequency images are fed into the multi-channel deep CNN model for automated fault detection. The proposed multi-channel deep CNN model is multi-headed, considering the time-frequency domain information of each channel time-series data for automated fault detection. The proposed model's performance is compared to benchmark models regarding testing accuracy, total parameters, and model size. Experiments have shown that the proposed model outperforms benchmark models regarding classification accuracy. The proposed multi-channel CNN model has obtained the accuracy and F1-score values of 99.48% and 99% for fault classification using time-frequency images of multi-sensor data. Finally, the proposed model's performance is measured regarding inference time when deployed on edge computing devices such as the Raspberry Pi and the Nvidia Jetson AGX Xavier.

INDEX TERMS Deep CNN, machinery failure, Nvidia Jetson AGX Xavier, Raspberry Pi, time-frequency analysis, wavelet synchro-squeezed transform.

I. INTRODUCTION

The current machinery systems, such as chemical processes, vehicle dynamics, aero engines, electric machines, manufacturing systems, wind energy conversion systems, induction motors, and power networks, need safety due to different types of faults. The demand for the reliability and safety of the industries is increasing effectively [1]. A minor fault may lead to complete machinery failure and cause loss for the industry. Hence, this leads to catastrophic accidents and destructive damage. Further, modern industry primarily uses rotating machinery. Under severe working conditions, there is more possibility for critical component failure due to high speed, big load, and heavy background noise [2], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain.

It is of great interest to reduce accidents and improve the reliability and safety of mechanical equipment. The multi-sensor-based systems are widely used to collect vibration data from the machine under normal and faulty conditions [4]. The time-series data recorded using these sensor systems have shown different characteristics for normal and faulty operations of the machine. The advances in artificial intelligence (AI)-based techniques enable the automated recognition of machine faults using multi-sensor vibration time-series data [5]. The development of novel AI-based methods using multi-senor data is vital for automated fault identification at the early stage. Motivated by this, a multi-channel deep convolutional neural network model is proposed for machinery fault diagnosis in this work. The following are the significant contributions of the proposed work:

- The multi-channel time-frequency domain deep CNN model is proposed for machine fault recognition using Multi-sensor data.
- Wavelet synchro-squeezed transform (WSST) is employed for the time-frequency analysis of the multi-sensor data.
- A multi-channel deep convolutional neural network model is developed to process the time-frequency images of sensor time-series data simultaneously.
- The performance of the proposed model is compared to the benchmark models through extensive results.
- Finally, the performance of the proposed model is evaluated in terms of inference time when deployed on edge computing devices such as Raspberry Pi and Nvidia Jetson AGX Xavier.

The paper is further organized as follows. Section III discusses the dataset details and the preprocessing with the WSST. Section IV describes the proposed CNN model architecture and the edge computing devices used. Section V presents the comparison results of the proposed model with the benchmark models in terms of accuracy and inference time. Finally, the conclusion and possible future works are discussed in Section VI.

II. RELATED WORK

A pattern recognition technique-based machinery fault diagnosis has been presented in [6] that deals with complicated signatures in vibration signals of rolling element bearings with or without defects. Further, it has been shown that the proposed method in [6] is very effective in terms of effective feature extraction, ability to learn, efficient feature fusion, and a simple algorithm for classification. Finally, their method's effectiveness has been tested with an experimental dataset. In [7], the authors have proposed a method to represent the rotating machinery fault diagnosis with the wavelet packet basis. The signal vector is divided into two subspaces by representing them differently. Among them, one subspace contains the transient components excited by local defects, and the remaining components are contained in another. Further, a basis selection algorithm and Coifman and Wickerhauser's best basis algorithm have been considered for the former subspace and second subspace, respectively, for enhancing the detection of detects. Finally, their suggested method has been tested using the gearbox dataset and a rolling element bearing. Through results, it has been shown that the proposed method does not require any training samples [7]. A local mean decomposition (LMD) based rotating fault diagnosis has been proposed in [8]. LMD is a novel method used for time-frequency analysis that is suitable for processing amplitude and frequency-modulated signals. The proposed LMD approach has been compared with the empirical mode decomposition (EMD) method. The results show that the proposed LMD method outperforms the EMD [8]. Finally, the authors have applied LMD for the fault diagnosis of gear and roller bearings. It has been shown through analytical results from practical gearbox vibration signals that the LMD-based approach is effective in diagnosing the condition of gear and roller bearing accurately. In [9], the rolling element bearing faults have been represented using frequency spectra, and then the fuzzy logic techniques (FLTs) have been applied for classification. FLTs generate fuzzy numbers that give the similarity between the frequency spectra. It has been observed that selecting the correct combination of the fuzzy numbers leads to the correct classification of the faults [9]. The above-discussed methods are conventional approaches that have utilized time-frequency analysis to diagnose machinery faults. Next, we discuss the state-of-the-art artificial intelligence (AI)/machine learning (ML)-based machinery fault diagnosis.

Advances in AI and computer technology can be used to diagnose machinery faults. Many techniques, including support vector machines (SVMs) [10], self-organizing feature maps [11], expert systems, rough sets, fuzzy logic, and neural network models have been utilized for machinery fault diagnosis. A survey on AI-based methods has been presented in [5] for rotating machinery fault diagnosis, focusing on both theoretical background and industrial applications. Further, the survey in [5] briefed some AI algorithms' advantages, limitations, and new research trends. A survey on deep learning (DL) models has been presented in [12] for intelligent machinery fault diagnosis, including their achievements, challenges, and pros and cons for the future research scope. Due to the development of intelligent fault diagnosis methods using DL, CNN has gained attention for machinery fault diagnosis. In light of this, a survey on the CNN-based rotary machine fault diagnosis methods has been presented in [13]. In [13], the authors have initially described the necessity of the data preprocessing techniques, which are crucial for CNN for the reduction in the difficulty of the feature learning and improving the accuracy. Then, a review of the preprocessing methods is presented. Thereafter, an analysis has been presented on the main methods used in CNN-based intelligence systems [13]. Next, we explain in detail each AI/ML algorithm.

A new framework based on deep convolutional variable-beta variational autoencoder (VAE) has been proposed in [14]

for machinery fault diagnosis. The proposed scheme in [14] is an intelligent scheme that is used to extract the discriminative features. To reduce the number of data points and extract 2D data, a novel min-max algorithm and random sampling technique, respectively, have been proposed. It has been mentioned that this framework is a single framework that combines both preprocessing and classification. When tested on the case Western Reserve University vibration dataset (VD) and air compressor acoustic dataset, the proposed approach in [14] resulted in 99.93% and 99.91% accuracy, respectively. A machine fault diagnosis method based on a stacked sparse autoencoder (SAE) has been proposed in [15]. The essential information can be mined using the penalty term of the SAE, which further removes redundant information. Empirical mode decomposition and autoregressive (AR) models are used to preprocess the collected non-stationary and transient signals, extracting AR parameters based on the intrinsic mode functions (IMFs). The superiority of the proposed method has been validated with experiments with a test accuracy of 99.83% [15]. For diagnosing the machinery fault, a deep contractive auto-encoding network (DCAEN) model has been presented in [16]. The DCAEN model is constructed by unsupervised learning named contractive auto-encoder (CAE), which helps extract the features unsupervised. Further, the network removes the redundancy by adding a sparsity constraint to the loss function of CAE. Moreover, the DCAEN model can extract the features automatically from the raw data without creating artificial features. Finally, the proposed model in [16] has been validated through experiments that resulted in an accuracy of 99.60% on the rolling bearing dataset, which is higher than the state-of-the-art methods. In [17], fault diagnosis of rotating machinery has been made by developing an enhancement deep feature fusion method. First, the feature learning ability has been enhanced by constructing a new deep auto-encoder with denoising and contractive auto-encoders. Then, the quality of the features was improved by adopting locality-preserving projection. Finally, Softmax has been trained with the fused deep features for intelligent diagnosis. It has been shown through the results that the proposed model in [17] achieves an accuracy of 90-100%. A novel deep autoencoder feature learning method has been developed in [18] for rotating machinery fault diagnosis. The developed method in [18] first adopts the maximum correntropy for designing the loss function for the new deep autoencoder to enhance the feature learning from the obtained vibration signals. After that, the artificial fish swarm algorithm was utilized for optimizing the parameters of the deep autoencoder in adaptation to the signal features. Finally, the proposed method has achieved a testing accuracy of 87.8% with a computation time of 371.16 seconds. The above methods are based on the autoencoder and are more beneficial in extracting useful information from the input avoiding redundancy. Next, we discuss some of the state-of-the-art deep learning models for fault diagnosis in the machinery.

The least squares support vector machine (LSSVM) and deep structure-based fault diagnosis method have been proposed in [19] for rotary machine fault diagnosis. The proposed method learns the deep features layer by layer, using SVM in each layer, until better classification results are obtained. To reduce the time complexity associated with multiple layers of SVM, a deep sparse least squares support vector machine (DLSSVM) has been proposed that combines the sparse theory with the LSSVM. At last, the proposed DLSSVM has been compared to the existing methods for diagnosing centrifugal pump faults and rolling bearing faults. It has been shown through results that the proposed DLSSVM results in an accuracy of 97.75%, which is higher than the state-of-the-art methods [19]. For fault diagnosis in rotating machinery, a novel deep learning method has been proposed in [20]. Due to the difficulty in obtaining the labeled data from the machinery fault, the authors have considered data augmentation for creating additional samples for model training. With two augmentation methods and five augmentation techniques, the proposed method in [20] has achieved an accuracy of 99.9%. A deep transfer learning model has been proposed in [21] for the machinery fault classification with a minimal dataset. The proposed model simultaneously conducts the supervised classification and multiple adversarial domain adaptation to improve performance. It has been shown through the experiments that the proposed method outperforms other models considered in [21]. Next, we discuss the existing CNN-based models for machinery fault diagnosis.

Two feature extraction methods, such as CNN and bag-of-visual-word (BoVW), have been proposed in [22] for the feature extraction and classification of the rotating machinery conditions using Infrared (IR) Images. It has been shown through experiments that CNN is performing better than BoVW on the IR image data [22]. The usage of the Predictive Maintenance model with Convolutional Neural Network (PdM-CNN) has been presented in [23] to notify maintenance when there is a rotatory equipment fault. The proposed model in [23] uses the data from only one vibrating sensor installed on the motor-drive end bearing. It has been shown that PdM-CNN achieves an accuracy of 99.58% and 97.3% with two publicly available datasets. A fault diagnosis approach that combines the one-dimensional CNN, Gated Recurrent Unit (GRU), attention mechanism (AM), and knowledge graph has been proposed in [24] for mechanical equipment. In the proposed approach, the CNN model is used for feature extraction, and GRU and AM mechanisms are used for accurate feature extraction. An efficient bearing fault diagnosis scheme that uses the CNN and energy distribution maps (EDMs) of AE has been proposed in [25]. In this scheme, CNN automatically extracts the features from EDM, and then an ensemble classifier which is a combination of CNN and SVM, has been used for bearing fault diagnosis [25].

The above methods have not focused on processing multiple sensor data simultaneously, which results in a fast

fault diagnosis. Further, the machine learning models are not lightweight, making it difficult to deploy them on low-computing devices. Thus, a multi-channel deep convolutional neural network model is proposed for machinery fault diagnosis in this work which is lightweight.

III. DATASET AND PRE-PROCESSING

The Wavelet synchrosqueezed transform (WSST) is a time-frequency analysis used for analyzing the signals. The input signal for the WSST is a vector or a set of real values. The WSST algorithm decomposes the signal into components with time-varying frequency.

A. DATASET DETAILS

The database is collected by using multiple sensors deployed on SpectraQuest's Machinery Fault Simulator (MFS) Alignment-Balance-Vibration (ABVT). The multivariate time-series dataset consists of 1951 data files involving six specific simulated states such as normal function, imbalance fault, horizontal and vertical misalignment faults, and inner and outer bearing faults. The data was generated at the rate of 50kHz for 5 seconds. Each file includes eight columns of Tachometer signal, underhang, and overhang bearing accelerometer with three dimensions axial, radial, and tangential direction, and microphone data which collect noise information. The six simulated states of the dataset are

1) NORMAL SEQUENCES

Normal Sequences indicate the data with no fault. The 49 measurements of the normal sequence cover a range of rotation speeds in the range 737–3686 rpm with steps of approximately 60 rpm.

2) IMBALANCE FAULT

The imbalance fault class is simulated with different load values ranging from 6-35 grams. The rotational frequency is frozen at 49 for the normal operation case with a load of less than 30 grams. Meanwhile, the loads above 29 grams provide vibrations which are very impractical to achieve 3300 rpm of rotational frequency. This reason limits the number of distinct rotations.

3) HORIZONTAL MISALIGNMENT FAULTS

The Horizontal misalignment simulated data is generated by shifting the motor shaft by 0.5 mm, 1.0 mm, 1.5 mm, and 2.0 mm into MFS. The rotation frequency for each horizontal shift is in the same range as the normal operation case.

4) VERTICAL MISALIGNMENT FAULTS

Vertical misalignment of the motor shaft was induced by shifting the shaft horizontally by 0.51 mm, 0.63 mm, 1.27 mm, 1.40 mm, 17.8 mm, and 1.90 mm into MFS. For each vertical shift, the rotation frequency range is the same as the normal operation case.

5) BEARING FAULTS

Rolling bearings are one of the most complex elements and susceptible elements to fault occurrence. The manufacturer of the ABVT provided three defectives with distinct defects—outer tracks, rolling elements, and inner tracks that were placed at different positions in the MFS experimental stand. Underhang: In MFS, the bearing is placed between the rotor and the motor. Overhang: In MFS, the rotor between the bearing and the motor. Fig. 1 shows the time series samples corresponding to different sensor data.

B. WAVELET SYNCHRO-SQUEEZED TRANSFORM

In this work, wavelet synchro-squeezed transform (WSST), which is an improved version of continuous wavelet transform (CWT) [26], is considered for the time-frequency analysis of the multi-sensor time-series data for the automated recognition of machine faults. WSST aims at sharpening the time-frequency graph by eliminating the smearing around the frequency values. In this study, we have not applied any filtering technique to remove the noise from the multi-sensor time-series data. The WSST approach acts as a filter for better localization of the signal components in both time and frequency domains of multi-sensor data. The deep learning models can be applied directly to the multi-sensor time-series to remove noise [27], [28].

Initially, WSST assumes that the signal can be represented as a linear combination of harmonics and some additive noise which is defined as [29] and [30]

$$s(t) = \sum_{l=1}^{L} A_l(t)\cos(\psi_l(t)) + n(t), \tag{1}$$

where, L denotes the number of signal components, A_l denotes the amplitude and ψ_l represents the phase of the l^{th} frequency component of the signal. $n(t)$ denotes the additive noise component at time instant t [31]. The instantaneous frequency (IF) for each component is obtained as

$$f_l(t) = \frac{1}{2\pi}\frac{d\psi_l(t)}{dt}. \tag{2}$$

Since most of the smearing happens along the frequency axis, the first step of WSST is to obtain the IF, $\omega_s(u, v)$, at each point (a,b). Here, $\omega_s(u, v)$ can be directly obtained from the CWT time-frequency representation, $W_s(u, v)$. Here, $\omega_s(u, v)$ is defined as

$$\omega_s(u, v) = \frac{-j}{2\pi W_s(u, v)}\frac{\partial W_S(u, v)}{\partial v}, \tag{3}$$

where j represents the imaginary number. Then, the smeared energy is relocated on these IFs by mapping the time-scale plane to the time-frequency, which is called synchrosqueezing. Finally, WSST can be obtained by

$$T_s(w_l, v) = \frac{1}{\Delta\omega}\sum_{u_k}\frac{W_S(u_k, v)}{\sqrt[3]{u}}\Delta u_k, \tag{4}$$

where, the scale, u, and time, b, are discrete values and $\Delta u_k = u_{k-1} - u_k$ is the scaling step [26]. Fig. 2 shows the

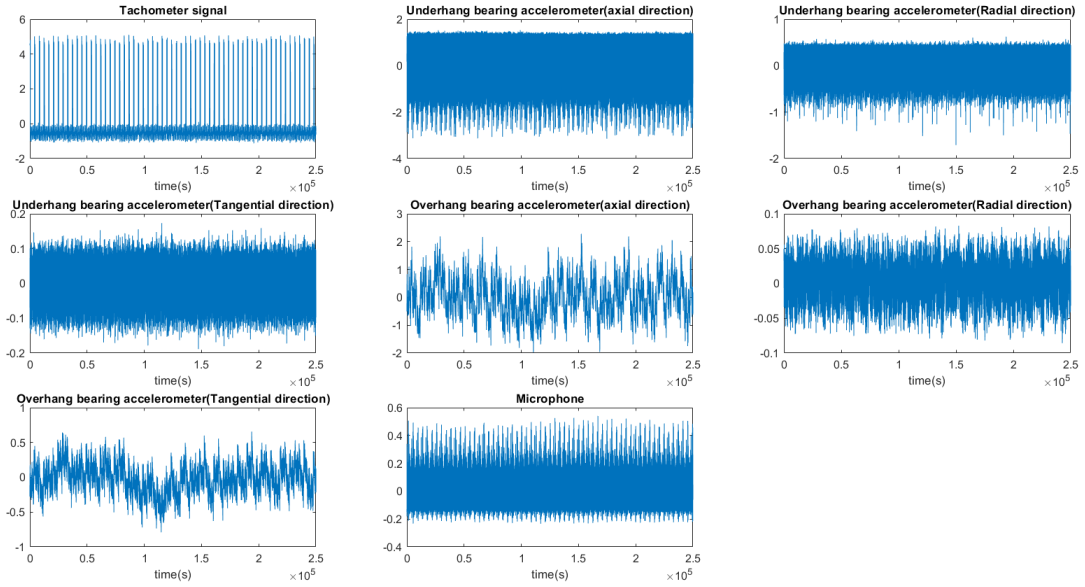

FIGURE 1. An illustration of oscillogram images of raw data of the samples.

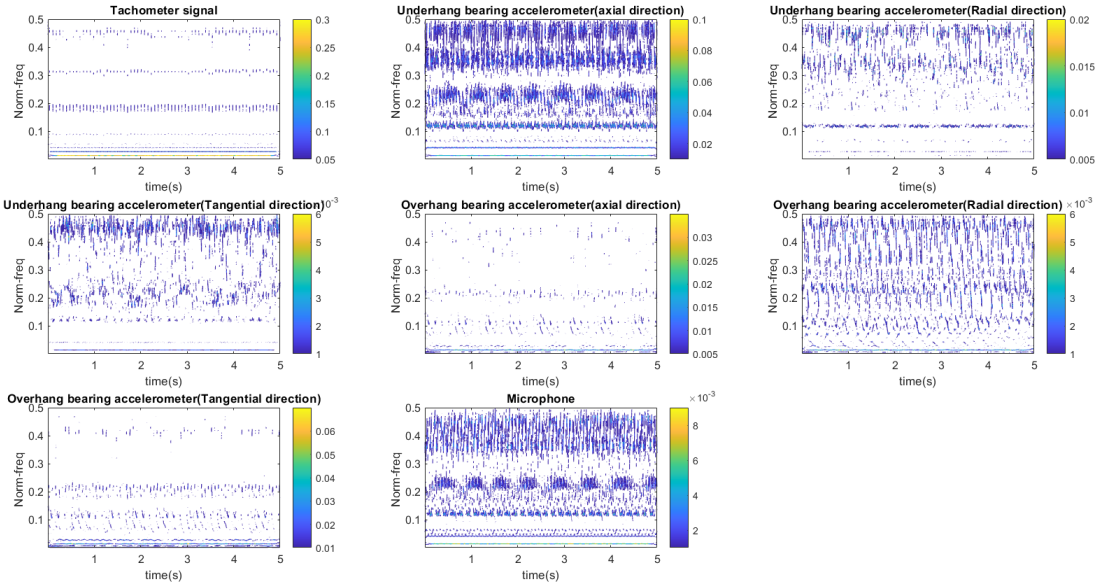

FIGURE 2. An illustration of the wavelet synchrosqueezed transform of the raw data samples.

time-frequency plots corresponding to the time series data in Fig. 1.

IV. PROPOSED CONVOLUTIONAL NEURAL NETWORK

Fig. 3 shows the architecture of the proposed multichannel Convolutional neural network model. Here, we consider feature-level fusion wherein the data features obtained from the eight channels are concatenated along a common axis to create a single, merged feature vector. This merged feature vector is then processed using a classifier, as can be seen from Fig. 3. The layer structure of all channels is the same, with each channel consisting of a batch normalization layer, convolution layer, max-pooling layer, convolution layer, max-pooling layer, convolution layer, max-pooling layers, and flatten layer. The outputs of flatten layers of all channels are combined by concatenation, and after that softmax layer is provided to give the output class as shown in Fig. 3.

Batch Normalization Layer: The Batch Normalization layer standardizes and normalizes the output layer based on the input layer. It helps the neural networks to become faster and more stable.

Convolution Layer: The convolution layer extracts the learnable features by applying a filter or kernel on the input layer. It creates a features map that consists of the summary of

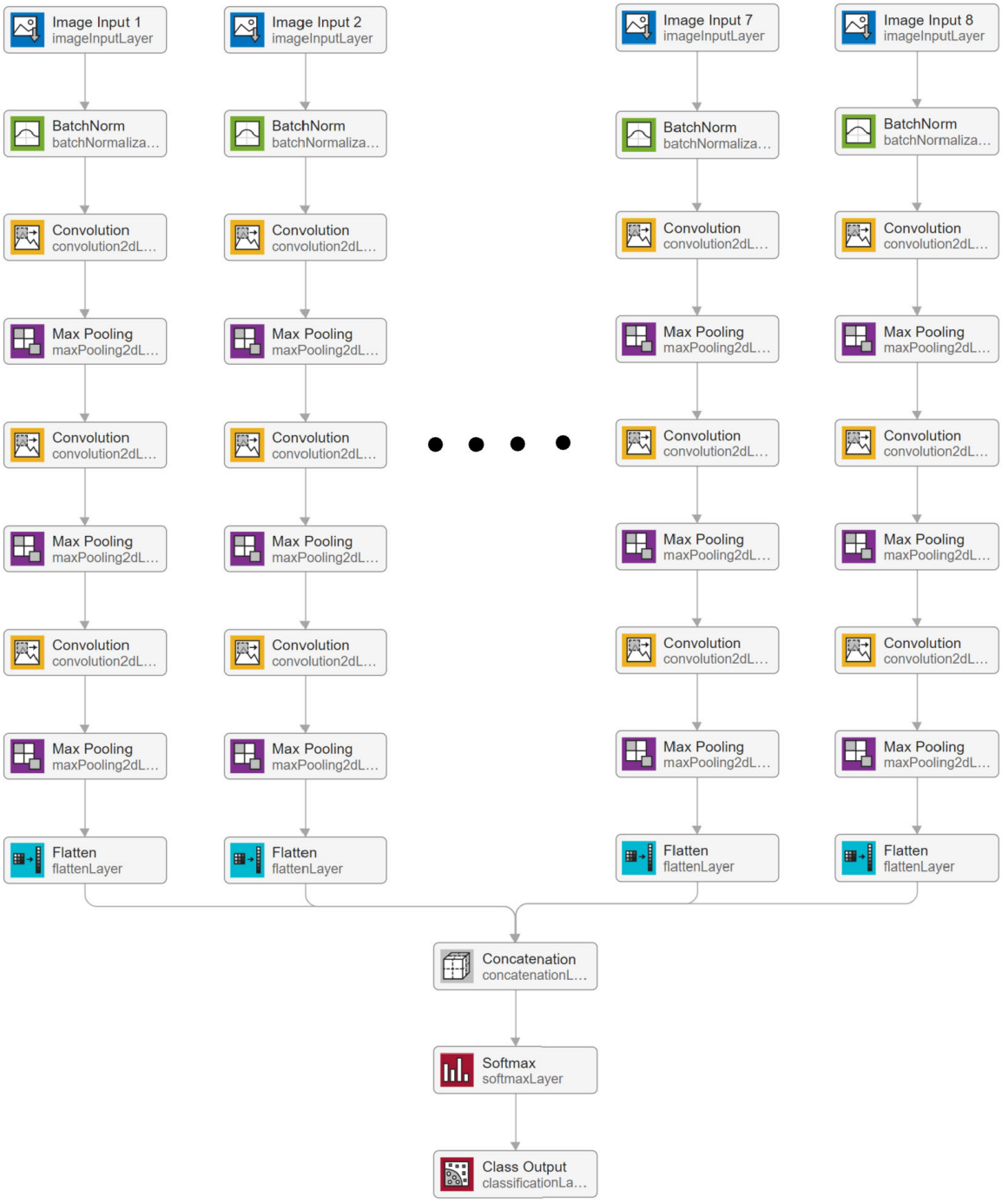

FIGURE 3. The proposed multi-channel deep CNN architecture.

the described features on the input. The convolutional layer output, O(x,y), can be calculated as

$$O(x, y) = (a * b)[x, y] = \sum_{n} \sum_{m} b[n, m]a[x - n, y - m], \tag{5}$$

where the input image is represented by a, and b denotes the kernel. The indices of rows and columns of the output matrix are represented by n and m.

Tanh Activation Function: It is a zero-centered function. This function output values are either -1 or 1 based on the real input value. The larger inputs and smaller inputs will be mapped strongly positive and strongly negative, respectively.

Max-Pooling Layer: Pooling layers are used to reduce the dimensionality of the input matrix. In the Max pooling layer, the kernel pulls the maximum value from the area of the feature map. The output of the pooling layer can be

TABLE 1. Number of samples per class in the dataset.

Class Name	Number of Images
horizontal-misalignment	197
imbalance	333
normal	49
overhang	513
underhang	558
vertical-misalignment	301
Total	**1951**

calculated as

$$\text{Output Size} = \frac{i - k + 2p}{s} + 1\,, \tag{6}$$

where, 'i' represents input length, 'k' denotes kernel size. 'p' and 's' denote the sizes of padding and stride, respectively.

Fully Connected Layer: It is the last layer in the Neural Network. It will form the final output by compiling the data extracted from the previous layers.

Softmax Activation Function: This function takes the vector k real values as input and results in an output of real values whose sum is 1. The standard softmax function is defined as

$$\text{Softmax layer } (i_j) = \frac{e^i_j}{\sum_{m=1}^{N} e^i_m} \tag{7}$$

Cross-Entropy Layer: This is the layer used to measure the performance of the model. The loss can be calculated as

$$loss = -\frac{1}{x}\sum_{a=1}^{x}\sum_{b=1}^{K}(G_{i,j}.ln(H_{i,j})), \tag{8}$$

where x denotes the number of training samples and K denotes the number of classes. $G_{i,j}$ and $H_{i,j}$ denote the actual output and calculated hypothesis of the network, respectively.

The size of the time-frequency image for each channel is 224 × 224. The proposed deep CNN model is multi-headed, and each head takes the input as the time-frequency image of one channel of multivariate time series data. Similarly, the deep CNN model's output has six neurons corresponding to 6 classes (5 faults and 1 normal). This work evaluates 1951 multichannel time-frequency images for different machinery faults and normal classes. The number of instances or multichannel time-frequency images from each class is shown in Table 1. The dataset is divided into three parts: training, validation, and testing for hold-out validation. We consider 10% of the instances for testing the model. In the rest, 90% of the instances are considered for training and 10% for validation of the proposed deep CNN model. For 10-fold cross-validation, each fold selects the training and test instances from different portions of the dataset. Like in the first fold, the initial 90% of the instances (index 1 to index 1755) or multichannel time-frequency images are used to train the deep CNN model, and the last 10% instances are employed for the testing phase of the model for the fault recognition.

A. HARDWARE DEPLOYMENT

We consider edge computing devices such as Raspberry Pi 4 Model B and Nvidia Jetson AGX Xavier to evaluate the performance of all models considered regarding inference time.

Raspberry Pi 4 Model B, as shown in Fig. 4, is a mini computer that is faster, more powerful, re-engineered, and completely upgraded than previous models. It has an ARM7 quad-core processor with 1.5 GHz speed and supports up to 8 GB SDRAM. It has two USB 3.0 and two USB 2.0 ports. It holds Ethernet, Wi-Fi, and Bluetooth with both 2.4 GHz and 5GHz operating frequencies. It also has a camera, audio, and composite video ports.

Nvidia Jetson AGX Xavier is built on a 512-core Volta GPU with Tensor Cores, along with an 8-core ARM v8.2 64-bit CPU, with 8MB L2 and 4MB L3 as shown in Fig. 4. It has a large memory of 32GB RAM, which is 256-bit LPDDR4x, and internal storage of 32GB, type eMMC 5.1. It is installed with a 7-way VLIW Vision Processor Accelerator as well as two NVDLA Engines for DL Acceleration. The physical measurements are $105mm \times 105mm \times 65mm$ for the entire module.

We also consider other hardware models such as GPU: P100-PCIE-16GB, GPU: Tesla T4, GPU: GTX 1050 4GB, GPU: Tesla K80,CPU: Intel(R) Xeon(R) Platinum 8259CL CPU @ 2.50GHz,CPU: Intel i5 8th gen, CPU: Intel(R) Xeon(R) Platinum 8168 @ 2.70GHz, and CPU: Intel i9 11th gen.

V. RESULTS AND DISCUSSION

The proposed eight-channel CNN model is tested on the fault machinery dataset with six classes: normal, imbalance fault, horizontal misalignment, vertical misalignment, inner bearing faults, and outer bearing faults. The performance improvement of the proposed architecture is due to the enabling of multi-channel processing for various sensor time series features.

Fig. 5 and Fig. 6 show the variation of accuracy and loss with epoch, respectively. It is observed from Fig. 5 that the training accuracy and validation accuracy reached almost 100% after eight epochs. Further, it is observed from Fig. 6 that the training and validation loss reach almost zero after epoch nine. Fig. 7 provides the confusion matrix corresponding to all six classes. It is observed from Fig. 7 that the horizontal misalignment is 0.95 confidence in itself and 0.05 with the imbalance class. Meanwhile, the rest of the classes achieved 100% accuracy.

The performance of the proposed model is evaluated when tested on individual sensor datasets as well as all sensor datasets in terms of the achieved test accuracy, parameters, and model size Table 2. It is observed from Table 2 that the proposed model achieves an accuracy of 99.48% when processing the data of all sensors simultaneously with total parameters of 68,742, and it consumes only 1.1MB. Further, the proposed model using 10-fold cross-validation (CV) gave an average classification accuracy of 99.28%. It is

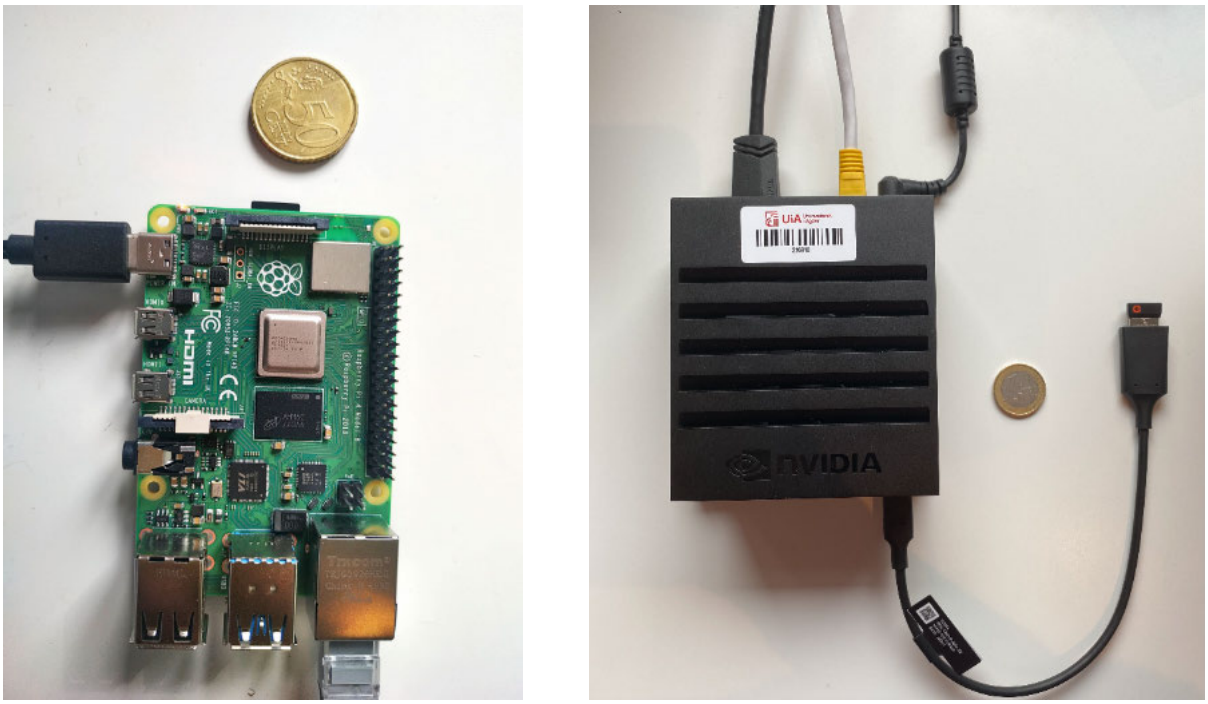

FIGURE 4. (a) Raspberry Pi 4 model B and (b) Jetson AGX Xavier.

TABLE 2. Test accuracy, total parameters, trainable parameters, non-trainable parameters, and size for the proposed model on multisensory data.

Data	Test accuracy	Total parameters	Trainable Parameters	Non-trainable Parameters	Model size
Tachometer Signal	28.06%	8,598	8,592	6	157 KB
Microphone	69.89%	8,598	8,592	6	157 KB
underhang bearing accelerometer	98.46%	25,782	25,764	18	441 KB
overhang bearing accelerometer	97.95%	25,782	25,764	18	441 KB
Combined Data	99.48%	68,742	68,694	48	1.1 MB

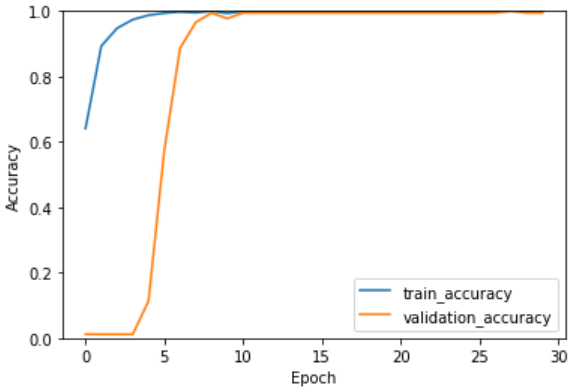

FIGURE 5. Variation of the accuracy of the proposed model with epoch.

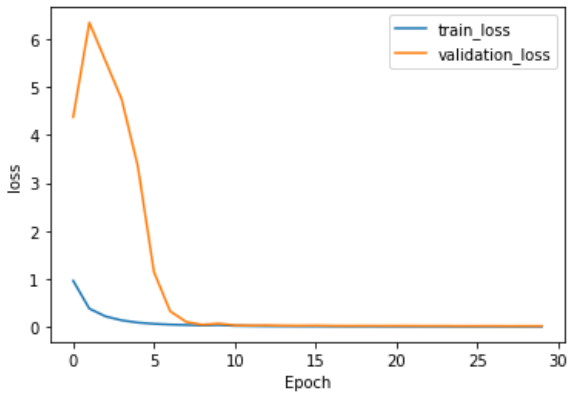

FIGURE 6. Variation of the loss of the proposed model with epoch.

also observed that the proposed model results in higher test accuracy when we process the data from all sensors. Further, the model size is light, which helps in real-time deployment. However, the proposed model performance degrades when considering data from one sensor alone.

Interestingly, the overall size of the proposed multi-channel architecture is still realizable with less size compared to benchmark models, as can be observed from Table 3. Table 3 provides the validation accuracy, total parameter, and size for all the models considered in this work. The

TABLE 3. The performance comparison of the proposed model with benchmark models.

Model	Testing Accuracy (10-fold)	Testing Accuracy (Holdout - 10%)	Total Parameters	Trainable Parameters	Non-trainable Parameters	Size
Proposed Model	66.53(±3.0)	65.81%	8,598	8,592	6	157 KB
DenseNet121	43.41(±2.48)	42.34%	7,043,654	6,150	7,037,504	27.6 MB
DenseNet169	50.71(±1.77)	46.93%	12,652,870	9,990	12,642,880	49.3 MB
DenseNet201	47.44(±1.22)	47.44%	18,333,510	11,526	18,321,984	71.1 MB
InceptionResNetV2	46.12(±2.06)	41.32%	54,345,958	9,222	54,336,736	208.6 MB
InceptionV3	41.42(±1.49)	47.44%	21,815,078	12,294	21,802,784	83.8 MB
MobileNetV2	54.69(±2.84)	51.53%	2,265,670	7,686	2,257,984	9 MB
MobileNetV3Large	41.58(±2.55)	41.32%	3,002,118	5,766	2,996,352	11.8 MB
MobileNetV3Small	45.66(±2.27)	41.83%	942,582	3,462	939,120	3.9 MB
NASNetMobile	37.80(±2.09)	46.93%	4,276,058	6,342	4,269,716	17.6 MB
ResNet50V2	43.26(±2.17)	47.44%	23,577,094	12,294	23,564,800	90.3 MB
ResNet101V2	48.62(±1.98)	51.02%	42,638,854	12,294	42,626,560	163.3 MB
ResNet152V2	45(±3.02)	53.06%	58,343,942	12,294	58,331,648	223.6 MB
VGG16	47.90(±1.33)	53.57%	14,717,766	3,078	14,714,688	56.2 MB
VGG19	39.43(±1.97)	49.48%	20,027,462	3,078	20,024,384	76.5 MB
Xception	46.64(±1.10)	45.91%	20,873,774	12,294	20,861,480	80 MB
EfficientNetB0	46.63(±1.48)	40.3%	4,057,257	7,686	4,049,571	15.9 MB
EfficientNetB1	50.15(±2.16)	43.36%	6,582,925	7,686	6,575,239	25.7 MB
EfficientNetB2	40.66(±1.73)	47.44%	7,777,023	8,454	7,768,569	30.3 MB
EfficientNetB3	50.4(±1.36)	52.55%	10,792,757	9,222	10,783,535	41.8 MB
EfficientNetB4	44.84(±1.46)	42.85%	17,684,581	10,758	17,673,823	68.3 MB
EfficientNetB5	41.83(±3.54)	39.79%	28,525,821	12,294	28,513,527	109.8 MB

TABLE 4. The inference time for the proposed model on different hardware modules.

Model	Inference time(milli-seconds)						
	GPU: Tesla T4	GPU: NVIDIA GeForce GTX 1050	CPU: Intel Core i5 8th gen	CPU: Intel Core i7 9750H	Intel(R) Xeon(R) Platinum 8259CL CPU @ 2.50GHz	Raspberry Pi	AGX
Microphone	5.09	4.73	6.34	5.26	3.63	66.58	11.8
Tachometer Signal	5.07	4.65	5.49	5.24	3.62	76.22	11.93
underhang bearing accelerometer	159.71	206.47	335..37	280.38	185.69	3990.41	748.07
overhang bearing accelerometer	167.43	211.83	339.12	280.63	176.62	4205.20	557.8
Combined Data	105.58	213.75	352.53	261.89	164.39	4299.42	583.84

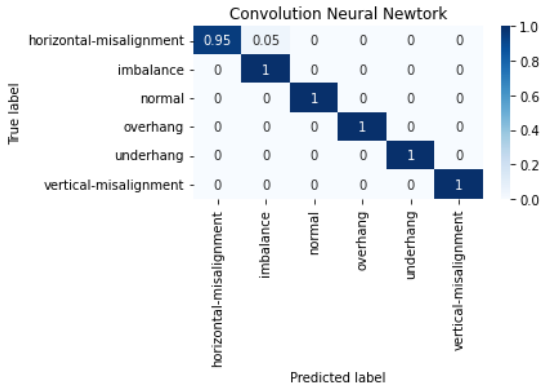

FIGURE 7. Confusion matrix of the proposed model.

proposed model is compared with the existing state of arts of DenseNet121, DenseNet169, DenseNet201, InceptionResNetV2, InceptionV3, MobileNetV2, MobileNetV3, MobileNetV3Large, MobileNetV3Small, NASNEtMobile, ResNet50V2, resNet101V2, ResNet152V2, VGG16, VGG19, Xception, EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4, and EfficientNetB5 when run on microphone data. This investigation uses the 10-fold validation technique to see the robustness of the proposed algorithm. It is evident from Table 3 that the proposed model gives superior performance compared to other models with 66.53% of ten-fold accuracy. InceptionV3 and EfficientNetB6 are bulk in size (> 100MB) due to many layers. Further, these bulk networks can only provide an accuracy of around 41%, much lower than the proposed architecture. The lightweight architecture and the small size of the proposed architecture can easily enable real-time applications like industrial automation, vital environment control, robotics, etc. It is also worth noting that the proposed architecture contains only 8598 parameters, 8592 trainable and 6 non-trainable parameters, with a 157KB size. The second best model, MobileNetV2, has a size of 9MB and provides 51.53% accuracy. At the same time, the proposed architecture can provide high performance and lightweight compared to the MobileNetV2, which is superior and highly applicable in industrial applications.

Table 4 shows the inference time required by the proposed model when deployed on edge computing devices. Here, we processed combined data from all the sensors as well as individual sensor data. It is observed from Table 4 that the microphone data and tachometer data require less

inference time. Whereas the data from the underhang bearing accelerometer and overhang bearing accelerometer requires higher inference time. When we consider the combined data, the inference time is greater than using microphone and tachometer data and less than underhang and overhang bearing accelerometer data, as can be seen from Table 4.

In this work, the number of test instances is less for the proposed time-frequency domain deep CNN model to recognize machinery faults. Access to large and diverse datasets is limited, especially in this specialized domain. The proposed study is useful for preliminary analysis, exploratory research, or proof of concept studies. This can help researchers assess the feasibility of a particular approach before investing in extensive data collection. This can serve as a starting point for refining approaches before applying them to larger datasets. The robustness of the model can be verified using new datasets with more instances to recognize more machinery faults. This work uses only four types of fault classes to develop the deep learning model. In the future, more number multivariate time series instances with different fault types can be utilized for the automated recognition of machinery faults.

VI. CONCLUSION

In this paper, a multi-channel deep CNN model for detecting machinery faults using multi-sensor time-series data has been proposed. The time-frequency images were obtained from the multi-sensor time series using the WSST-based time-frequency analysis technique. A deep CNN model considered time-frequency images of each channel time series to recognize different types of faults. Experiments have shown that the proposed multichannel deep CNN model achieves a 10-fold testing accuracy of 99.28($\pm$0.4) when processing the combined data from all sensors. We also compared the proposed model's performance to that of the benchmark models. Experiments have also shown that the proposed model achieves a 10-fold testing accuracy of 66.53($\pm$3.0) on the microphone dataset, which is higher than other benchmark models. Furthermore, it has been demonstrated that the proposed model is lightweight, with 8,598 parameters and a size of 157 KB, which is significantly less than other models. Finally, the proposed model's performance in terms of inference time has been compared to other models when deployed on edge computing devices such as the Raspberry Pi and Nvidia AGX Xavier. In the future, we plan to generate more samples to analyze the performance of the proposed model as compared to other models.

REFERENCES

[1] Z. Feng, D. Zhang, and M. J. Zuo, "Adaptive mode decomposition methods and their applications in signal analysis for machinery fault diagnosis: A review with examples," *IEEE Access*, vol. 5, pp. 24301–24331, 2017.

[2] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, and Z. He, "Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 1–35, Mar. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327015003854

[3] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review," *IEEE Access*, vol. 7, pp. 122644–122662, 2019.

[4] S. Li, Y. Xin, X. Li, J. Wang, and K. Xu, "A review on the signal processing methods of rotating machinery fault diagnosis," in *Proc. IEEE 8th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, May 2019, pp. 1559–1565.

[5] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.

[6] Q. Sun, P. Chen, D. Zhang, and F. Xi, "Pattern recognition for automatic machinery fault diagnosis," *J. Vib. Acoust.*, vol. 126, no. 2, pp. 307–316, Apr. 2004.

[7] B. Liu, "Selection of wavelet packet basis for rotating machinery fault diagnosis," *J. Sound Vib.*, vol. 284, nos. 3–5, pp. 567–582, Jun. 2005.

[8] J. Cheng, Y. Yang, and Y. Yang, "A rotating machinery fault diagnosis method based on local mean decomposition," *Digit. Signal Process.*, vol. 22, no. 2, pp. 356–366, Mar. 2012.

[9] C. K. Mechefske, "Objective machinery fault diagnosis using fuzzy logic," *Mech. Syst. Signal Process.*, vol. 12, no. 6, pp. 855–862, Nov. 1998.

[10] S.-F. Yuan and F.-L. Chu, "Fault diagnostics based on particle swarm optimisation and support vector machines," *Mech. Syst. Signal Process.*, vol. 21, no. 4, pp. 1787–1798, May 2007.

[11] T. Kohonen, *Self-Organization and Associative Memory*, vol. 8. Cham, Switzerland: Springer, 2012.

[12] C. Li, S. Zhang, Y. Qin, and E. Estupinan, "A systematic review of deep transfer learning for machinery fault diagnosis," *Neurocomputing*, vol. 407, pp. 121–135, Sep. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231220306123

[13] S. Tang, S. Yuan, and Y. Zhu, "Data preprocessing techniques in convolutional neural network based on fault diagnosis towards rotating machinery," *IEEE Access*, vol. 8, pp. 149487–149496, 2020.

[14] G. Dewangan and S. Maurya, "Fault diagnosis of machines using deep convolutional beta-variational autoencoder," *IEEE Trans. Artif. Intell.*, vol. 3, no. 2, pp. 287–296, Apr. 2022.

[15] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017.

[16] J. Yi, S. Fu, S. Cui, and C. Zhao, "A deep contractive auto-encoding network for machinery fault diagnosis," in *Proc. 18th Int. Symp. Commun. Inf. Technol. (ISCIT)*, Sep. 2018, pp. 1–5.

[17] H. Shao, H. Jiang, F. Wang, and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 119, pp. 200–220, Mar. 2017.

[18] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 95, pp. 187–204, Oct. 2017.

[19] K. Li, R. Zhang, F. Li, L. Su, H. Wang, and P. Chen, "A new rotation machinery fault diagnosis method based on deep structure and sparse least squares support vector machine," *IEEE Access*, vol. 7, pp. 26571–26580, 2019.

[20] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation," *J. Intell. Manuf.*, vol. 31, no. 2, pp. 433–452, Feb. 2020.

[21] T. Han, C. Liu, R. Wu, and D. Jiang, "Deep transfer learning with limited data for machinery fault diagnosis," *Appl. Soft Comput.*, vol. 103, May 2021, Art. no. 107150.

[22] Z. Jia, Z. Liu, C.-M. Vong, and M. Pecht, "A rotating machinery fault diagnosis method based on feature learning of thermal images," *IEEE Access*, vol. 7, pp. 12348–12359, 2019.

[23] R. M. Souza, E. G. S. Nascimento, U. A. Miranda, W. J. D. Silva, and H. A. Lepikson, "Deep learning for diagnosis and classification of faults in industrial rotating machinery," *Comput. Ind. Eng.*, vol. 153, Mar. 2021, Art. no. 107060. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360835220307300

[24] H. Liu, R. Ma, D. Li, L. Yan, and Z. Ma, "Machinery fault diagnosis based on deep learning for time series analysis and knowledge graphs," *J. Signal Process. Syst.*, vol. 93, no. 12, pp. 1433–1455, Dec. 2021.

[25] V. Tra, S. A. Khan, and J.-M. Kim, "Diagnosis of bearing defects under variable speed conditions using energy distribution maps of acoustic emission spectra and convolutional neural networks," *J. Acoust. Soc. Amer.*, vol. 144, no. 4, pp. EL322–EL327, Oct. 2018.

[26] S. K. Ghosh, R. K. Tripathy, R. N. Ponnalagu, and R. B. Pachori, "Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features," *IEEE Sensors Lett.*, vol. 3, no. 12, pp. 1–4, Dec. 2019.
[27] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[28] D.-G. Kim, Y. Ali, M. A. Farooq, A. Mushtaq, M. A. A. Rehman, and Z. H. Shamsi, "Hybrid deep learning framework for reduction of mixed noise via low rank noise estimation," *IEEE Access*, vol. 10, pp. 46738–46752, 2022.
[29] P. Wiciak, G. Cascante, and M. A. Polak, "Novel application of wavelet synchrosqueezed transform (WSST) in laser-vibrometer measurements for condition assessment of cementitious materials," *NDT & E Int.*, vol. 120, Jun. 2021, Art. no. 102424.
[30] J. B. Tary, R. H. Herrera, and M. van der Baan, "Analysis of time-varying signals using continuous wavelet and synchrosqueezed transforms," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 376, no. 2126, Aug. 2018, Art. no. 20170254.
[31] I. Daubechies, J. Lu, and H.-T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 243–261, Mar. 2011.

RAKESH REDDY YAKKATI received the M.Sc. degree in mathematics and computing from the Birla Institute of Technology, Mesra, Ranchi. His current research interests include artificial neural networks, machine learning, deep learning, and reinforcement Learning.

SREENIVASA REDDY YEDURI (Member, IEEE) received the bachelor's degree in electronics and communication engineering from Andhra University, Visakhapatnam, India, in 2013, the master's degree in computer science engineering (advanced networks) from the ABV-Indian Institute of Information Technology, Gwalior, India, in 2016, and the Ph.D. degree in electronics and communication engineering from the National Institute of Technology, Goa, India, in 2021. He has carried out the Ph.D. work with the Department of Electrical Engineering, Indian Institute of Technology Hyderabad, India. He is currently a Postdoctoral Research Fellow with the Autonomous and Cyber-Physical Systems (ACPS) Research Group, Department of ICT, University of Agder, Grimstad, Norway. His current research interests include AI/ML in communication/networking, cyber-physical systems, machine-type communications, the Internet of Things, LTE MAC, 5G MAC, optimization in communication, wireless networks, power line communications, visible light communications, hybrid communication systems, spectrum cartography, spectrum sensing, UAV-assisted wireless networks, wireless sensor networks, UAV path planning, and computer vision.

RAJESH KUMAR TRIPATHY (Member, IEEE) received the B.Tech. degree in electronics and telecommunication engineering from the Biju Patnaik University of Technology (BPUT), Odisha, India, in 2009, and the M.Tech. degree in biomedical engineering from the National Institute of Technology (NIT) Rourkela, Rourkela, India, in 2013, and the Ph.D. degree in machine learning for biomedical signal processing from the Electronics and Electrical Engineering (EEE) Department, Indian Institute of Technology (IIT) Guwahati, Guwahati, India, in 2017. He was an Assistant Professor with the Faculty of Engineering and Technology (FET), Siksha 'O' Anusandhan Deemed to be University, from March 2017 to June 2018. Since July 2018, he has been an Assistant Professor with the Department of Electrical and Electronics Engineering (EEE), Birla Institute of Technology and Science (BITS), Pilani, Hyderabad Campus. He has published research papers in reputed international journals and conferences. His current research interests include machine learning, deep learning, biomedical signal processing, sensor data processing, medical image processing, and the Internet of Things (IoT) for healthcare. He has served as a reviewer for more than 15 scientific journals and served as a technical program committee (TPC) member of various national and international conferences. He is an Associate Editor of IEEE Access and *Frontier in Physiology*.

LINGA REDDY CENKERAMADDI (Senior Member, IEEE) received the master's degree in electrical engineering from the Indian Institute of Technology Delhi (IIT Delhi), New Delhi, India, in 2004, and the Ph.D. degree in electrical engineering from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2011. Before joining the Ph.D. program with NTNU, he was with Texas Instruments, where he focused on mixed-signal circuit design. After finishing the Ph.D. degree, he was with the University of Bergen, Bergen, Norway, from 2010 to 2012, where he worked on radiation imaging for an atmosphere-space interaction monitor (ASIM Mission to the International Space Station). He is currently the Leader of the Autonomous and Cyber-Physical Systems (ACPS) Research Group and a Professor with the University of Agder, Grimstad, Norway. He is the Principal Investigator and a Co-Principal Investigator of many research grants from the Norwegian Research Council. His current research interests include the Internet of Things (IoT), cyber-physical systems, autonomous systems, robotics and automation involving advanced sensor systems, computer vision, thermal imaging, LiDAR imaging, radar imaging, wireless sensor networks, smart electronic systems, advanced machine learning techniques, and connected autonomous systems, including drones/unmanned aerial vehicles (UAVs), unmanned ground vehicles (UGVs), unmanned underwater systems (UUSs), 5G- (and beyond) enabled autonomous vehicles, and socio-technical systems, like urban transportation systems, smart agriculture, and smart cities. He is also quite active in medical imaging. He has coauthored over 150 research publications that have been published in prestigious international journals and standard conferences in the research areas. He is a member of ACM and the editorial boards of various international journals and the technical program committees of several IEEE conferences. Several of his master's students won the Best Master's Thesis Award in Information and Communication Technology (ICT). He serves as a reviewer for several reputed international conferences and IEEE journals.