A wavelet convolutional capsule network with modified super resolution generative adversarial network for fault diagnosis and classification

Article *in* Complex & Intelligent Systems · April 2022

DOI: 10.1007/s40747-022-00733-6

CITATIONS
14

READS
131

7 authors, including:

Happy Nkanta Monday
Oxford Brookes College of Chengdu University of Technology
55 PUBLICATIONS **596** CITATIONS
SEE PROFILE

Grace U. Nneji
Oxford Brookes College of Chengdu University of Technology
49 PUBLICATIONS **447** CITATIONS
SEE PROFILE

Saifun Nahar
University of Missouri - St. Louis
17 PUBLICATIONS **160** CITATIONS
SEE PROFILE

Md Altab Hossin
111 PUBLICATIONS **1,439** CITATIONS
SEE PROFILE

Complex & Intelligent Systems
https://doi.org/10.1007/s40747-022-00733-6

ORIGINAL ARTICLE

A wavelet convolutional capsule network with modified super resolution generative adversarial network for fault diagnosis and classification

Happy Nkanta Monday[1] · Jianping Li[1] · Grace Ugochi Nneji[2] · Saifun Nahar[3] · Md Altab Hossin[4] · Jehoiada Jackson[2] · Ariyo Oluwasanmi[2]

Received: 20 October 2021 / Accepted: 27 March 2022

Abstract
The study of fault diagnosis and classification has gained tremendous attention in various aspects of modern industry. However, the performance of traditional fault diagnosis technique solely depends on handcrafted features based on expert knowledge which is difficult to pre-design and has failed in several applications. Deep learning (DL) has achieved remarkable performance in hierarchical feature extraction and learning distinctive feature of dataset from related distribution. However, the challenge associated with DL models is that max-pooling operation usually leads to loss of spatial details during high-level feature extraction. Another concern is the low quality characteristics of 2D time-frequency image which is mostly caused by the presence of noise and poor resolution. This paper proposes a modified wavelet convolutional capsule network with modified enhanced super resolution generative adversarial network plus for fault diagnosis and classification. It uses continuous wavelet transform to convert raw data signals to 2D time-frequency images and applies super resolution generative adversarial technique to enhance the quality of the time-frequency images and finally, the convolutional capsule network learns the extracted high-level features without loss of spatial details for the diagnosis and classification of faults. We validated our proposed model on the famous motor bearing dataset from the Case Western Reserve University. The experimental results show that our proposed fault diagnostic model obtains higher diagnosis accuracy of 99.84% outweighing most traditional deep learning models including state-of-the-art methods.

Keywords Capsule network · CNN · Fault diagnosis · GAN · Wavelet · Super resolution

Introduction

In modern industry, one of the things that play a crucial role is fault diagnosis [1]. Data-driven fault diagnosis being a typical type of fault diagnosis has attracted much attention in recent

Jian Ping Li, Grace Ugochi Nneji, Saifun Nahar, Md Altab Hossin and Jehoiada Jackson have contributed equally to this work.

✉ Jianping Li
jpli2222@uestc.edu.cn

Happy Nkanta Monday
mh.nkanta@std.uestc.edu.cn

Grace Ugochi Nneji
ugochinneji@std.uestc.edu.cn

Saifun Nahar
snnnm@umsl.edu

Md Altab Hossin
altabbd@uestc.edu.cn

Jehoiada Jackson
kofijackson@uestc.edu.cn

Ariyo Oluwasanmi
ariyo@uestc.edu.cn

[1] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, China

[2] School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, China

[3] Department of Information System and Technology, University of Missouri St Louis, St. Louis, MO 63121, USA

[4] School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, China

Published online: 25 April 2022

years [2]. To carry out the extraction of the underlying knowledge concerning system variables, historical data of large volume is used especially for the complicated systems where it looks hard to establish unambiguous models or symptoms of signal [3]. The development of smart manufacturing has brought ease to the process of data collection [4] which does not only bring new perspective to the industry but has also brought some challenges [5]. Therefore, it becomes crucial to discover an effective data-driven method for fault diagnosis [6]. Traditional machine learning (ML) methods utilize predesigned handcrafted features and these features contribute to the best possible (upper bound) prediction accuracy [7]. In 2006, Deep Learning (DL) became the center of attraction for most researchers in the field of machine learning [8]. DL has the ability to automatically extract features of raw data in hierarchical representation [9,10]. This advantage enables the DL to avoid the errors encountered in the handcrafted features designed by domain experts and has consequently shown a remarkable prospect on the diagnosis of faults [11,12]. Although traditional machine learning methods perform well under the assumption that both the training and the testing data are expected to be drawn from exactly the same distribution. When drawn from differing distributions, then the performances would drop significantly. In addition to that, this assumption has never recorded success in many applications.

DL methods also encounter the same bottleneck mentioned above. To solve this challenge, a well-known approach called Transfer Learning (TL) method is utilized to perform learning task on both the training and testing dataset from a distribution that is related. TL approach offers better adaptability in extracting high-level features compared to shallow architectures and has recorded tremendous progress in many applications [13]. However, the challenges associated with TL models is that max-pooling operation usually leads to loss of spatial details during high-level feature extraction. According to Zellinger et al. [14] who suggested a unique strategy for unsupervised domain-adaptation for neural networks that depends on the regularization of the metric-based learning procedure. The authors further explained that by decreasing the suggested Central Moment Discrepancy (CMD) metric, the regularization tries to maximize the resemblance of domain-specific activation distributions. The authors also stated that the CMD addresses difficulties of instability that occur when using integral probability metrics based on polynomial function spaces. More so, the authors explained that in dual space, the metric can be interpreted as the sum of the differences between higher order central moments of the associated activation distributions [14]. Studies have shown that deep learning models combined with wavelet transform significantly increase the overall performance of the network in classification task compared to the traditional stand alone deep learning models [15]–[17].

This paper proposes a modified wavelet convolutional capsule network with modified enhanced super resolution generative adversarial network plus for fault diagnosis and classification. It uses continuous wavelet transform to convert raw data signals to 2D RGB images (scalograms) and applies super resolution generative adversarial technique to enhance the quality of the scalograms and finally, the convolutional capsule network learns the extracted high-level features without loss of spatial details for the diagnosis and classification of faults. We conducted several experiments to examine the diagnosis performance of our proposed MWCCN-MESRGAN+ model. From the results obtained, not only does MWCCN-MESRGAN+ has a promising potential in fault diagnosis, it also outweighs several TL models and fault diagnosis methods. The remaining part of this paper is structured as follows. The related works is presented in Sect. 2. Methodology and detailed discussion of the proposed framework are presented in Sect. 3. Section 4 presents the several experiments conducted. Section 5 presents the evaluation of the proposed framework and finally, the conclusion of this paper is given in Sect. 5.

Related works

Fault diagnosis abates the risk of unforeseen breakdown and ensures the safety as well as the reliability of industrial systems. Generally, the methods of fault diagnosis can be categorized into four domain subjects; signal domain, model domain, hybrid/active domain, and knowledge domain methods [18]. The knowledge domain method is data-driven technique which is better used for systems that are too complicated and difficult to obtain specific system framework or symptoms of signal [18]. Machine learning is a well-known analytical technique in data-driven fault diagnosis such as artificial neural network (ANN), expert system, Support Vector Machine (SVM) as well as fuzzy logic.

The first data-driven fault diagnosis that became famous was developed in the 1980s with the use of expert systems [19]. This method utilizes a technique that requires the expert to learn a set of rules from previous experiences. The authors in [20] suggested an extended version of neural network for fault diagnosis of internal combustion engines. A study was suggested in [21] to investigate the merits of SVM method for fault diagnosis and to monitor the recent progress. The authors in [22] proposed an intelligent framework which is based on a fuzzy genetic algorithm. This method was applied to automatically detect failures in aircraft.

The authors in [23] investigated the possible applications of DL to machine condition monitoring. Recurrent neural networks and dynamic Bayesian modeling approach was suggested by the authors in [24] to detect faults in induction motors. The investigation of an stacked Auto-Encoder

(SAE) for the classification of induction motor fault was proposed by the authors in [25]. Two-layer unsupervised neural network with sparse filtering technique was proposed in [26] for fault diagnosis. The authors in [27] applied the stacked de-noising auto-encoder (AE) to fault diagnosis in rotary machinery. The authors in [28] suggested an improved deep belief network (DBN) for fault diagnosis in rolling bearing. A deep neural network with auto-encoder (AE) was investigated by the authors in [29] for smart fault diagnosis.

Despite the huge achievement of machine learning methods,it still fails in many applications due to its assumption that the data distributions of the train and the test are the same [30]. The authors in [31] presented a study to analyze different methods while the work of [32] presented some of the applications of TL. A method of transfer component analysis was proposed by the authors in [33] to achieve feature transformation with the aim of discovering common latent features with similar margin of distribution while keeping the intrinsic structure of the input data.

A method of handling the heterogeneous features of both videos and images by transferring the SVM was proposed in [34]. A method of unsupervised Stacked de-noising AE was proposed in [35] to transform the input space to find consistent latent feature space. A study on transfer learning by reusing stacked de-noising AE was investigated by the authors in [36].The authors in [37] utilized domain adaption with neural network for fault diagnosis. It is worth mentioning that rolling bearing fault data are raw signals (time-series) data that need to be converted to 2D time-frequency images by means of continuous wavelet transform (CWT). These time-frequency images are absolute value of CWT coefficient characterized with low quality. However, all the methods mentioned above contributed remarkably to the study of fault diagnosis but none of the methods addressed the effect of low quality 2D time-frequency scalogram on the performance of fault diagnosis. The method proposed in this research is designed to solve the problem of low quality 2D time-frequency scalograms with the aim of obtaining high performance on fault diagnosis.

Methodology

This section will introduce the dataset utilized in this paper, the pre-processing of the fault dataset, the modified enhanced super resolution generative adversarial network plus, followed by the wavelet convolution capsule network, and the time-frequency scalogram construction. Finally, the experimental setup and details conclude this section.

Dataset

In this study, we collected raw signals of fault dataset from Case Western Reserve University [38] which consists of 10 health conditions with 1024 sample points each. Out of the 10 conditions, only 1 condition belongs to the normal label while the other 9 conditions are classified as fault condition with different damage points, fault diameters and load conditions. The raw signals of the 10 health conditions are sampled at the frequency of $12kHz$. For the purpose of our study, we split the dataset into train, validation and test.

Pre-processing of fault dataset

Dataset of raw fault signals are collected from a well-known bearing data repository of the Case Western Reserve University [38]. Since our proposed model requires the input data to be in image format, we converted the raw signals into time-frequency scalograms (images) using continues wavelet transform (CWT). However, the time-frequency scalograms are grayscale with 1 channel, therefore, it is important to convert them to 3 channel format of RGB. Finally, the input image is normalized and reshaped to a dimension of $224 \times 224 \times 3$ to match the input size of our proposed model. The dataset is subdivided into training, validation and test set.

Modified enhanced super resolution GAN Plus (MESRGAN+)

In this study, our aim is to enhance the low quality of scalogram (2D images) into a super-resolution before passing them through the wavelet convolutional capsule network for bearing fault diagnosis and classification. We will present the proposed modified enhanced super resolution generative adversarial network plus (MESRGAN+) architecture and its structural improvement for achieving a balance in perceptual quality and PSNR in this section. Hence, we will briefly highlight the transition of SRGAN to MESRGAN+.

Transition of super resolution by GAN

SRGAN [39] utilizes basic blocks of deep residual network to recover image-realistic details in which batch normalization(BN) is followed after each convolutional layer as depicted in Fig. 1. The transition from SRGAN to ESRGAN [40] is based on two modifications; the first modification is the removal of all BN in the generator structure and the second modification involves the replacement of the original basic block with Residual-in -Residual Dense Block (RRDB) as shown in Fig. 1. Finally, the transition from ESRGAN [40]

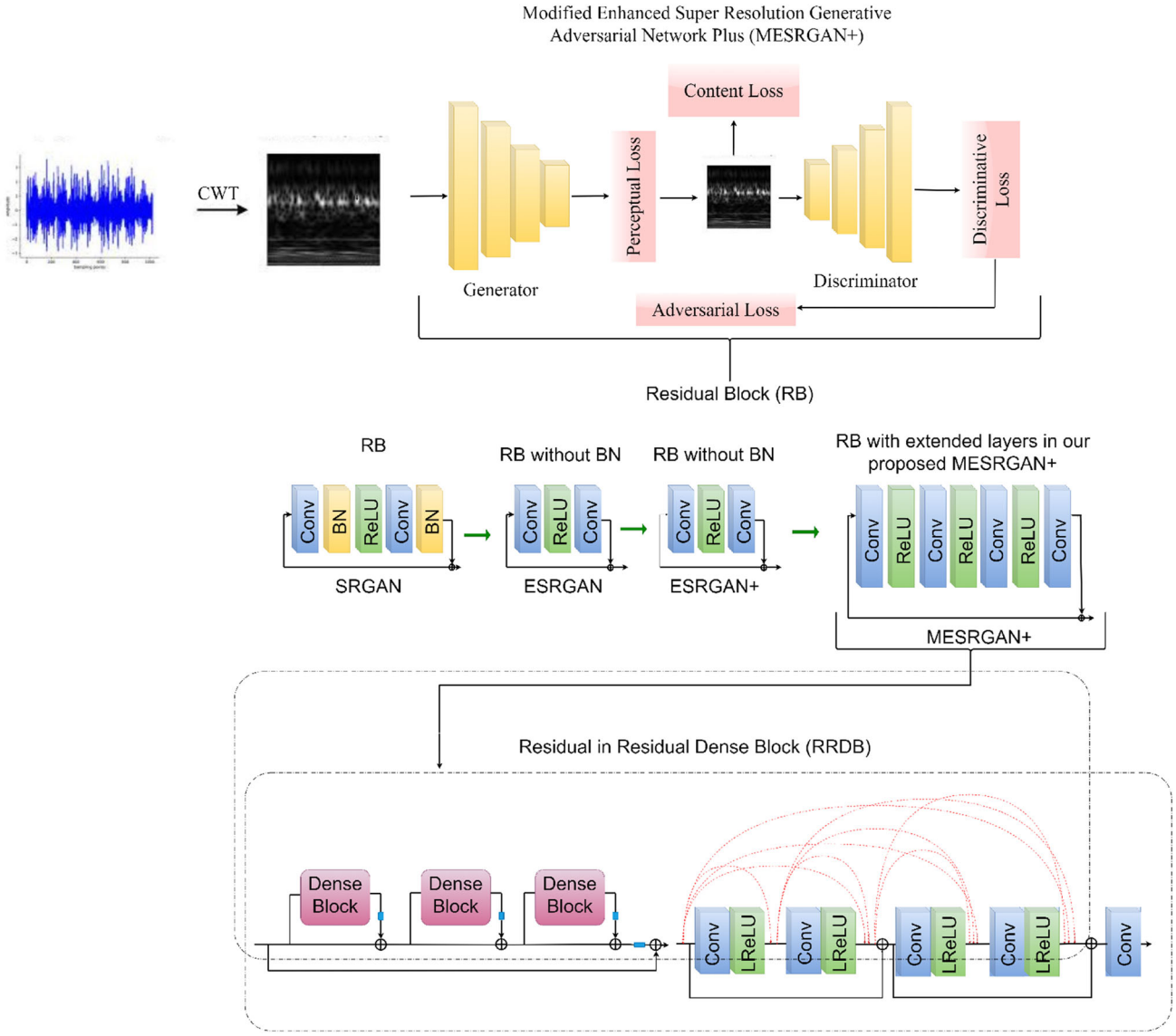

Fig. 1 Detailed structure of our proposed modified enhanced super resolution generative adversarial network plus (MESRGAN+)

to ESRGAN+ [41] is based on introducing additional level of residual learning at every two layers inside the dense block as illustrated in Fig. 1 without changing the convolutional structure.

Architecture of the proposed MESRGAN+

In our proposed super resolution architecture, the overall structural configuration of the Residual-in-Residual Dense Block (RRDB) in ESRGAN+ is kept the same as shown in Fig. 1. We made few modifications to the ESRGAN+ network in the generator structure by expanding the convolutional layers with additional two convolutional layers and two ReLU activation function. Normally, the direct mapping of the high-dimensional LR features to HR feature vectors ultimately results to high computational complexity and we know that the dimension of the LR feature is normally very huge. To address this bottleneck, we utilize a 1×1 convolutional layer as the second layer to reduce the computational cost by shrinking the LR dimensional features thereby maintaining the same kernel size of 64 after the first layer. To maintain consistency and the performance of ESRGAN+, we utilized 3×3 filter size and kernel size of 64 for the third and fourth convolutional layers.

To produce the high-resolution images from the scale-adaptive module, the scale factor is increased to 4. This image's network generator produces $v^{k+1} = G_k(v^k)$. Feature map is extracted to calculate the perceptual loss before being passed to the final activation function. Pixel-wise loss is measured, and the created image is forwarded to the discriminator network to differentiate between the created image v^{k+1} and the actual image $\hat{v}^{k+1}$. This actual image $\hat{v}^{k+1}$ is

fed to the discriminator network for training, which results in the same super-resolution image v^{k+1}. The generator network could generate new images that look like the real image. When training begins, the generator produces obviously fake data, and the discriminator quickly learns to tell that it's fake. As training progresses, the generator gets closer to producing output that can fool the discriminator. Finally, if the generator training goes well, the discriminator gets worse at telling the difference between real and fake. It starts to classify fake data as real, and its accuracy decreases. During the discriminator training, the discriminator classifies both real and fake image from the generator. Finally, if the generator training goes well, the discriminator gets worse to distinguish real and fake image which simply means that this entire process was only completed when the discriminator network could no longer tell the difference between real and fabricated images. At this point, the discriminator loss penalizes the discriminator for misclassifying a real instance as fake or a fake instance as real.Therefore, if the process is incomplete, the generator loss penalizes the generator for failing to fool the discriminator. At this point, the discriminator can still tell the difference between the fake and original image. We train the generator function G_k to approximate the HR of the next LR image $\hat{v}^{k+1}$ which LR input can represent.The total loss of the super-resolution network is given in (1) as;

$$\begin{aligned}\Pi_{Totalloss} =& \Pi_{Gen}\left(\Pi_{Perceptualloss} + \mu\Pi_{G}^{Ra} + \eta L1\right) \\ &+ \Pi_{Dis}^{Ra}\end{aligned} \tag{1}$$

As (1) is evaluated, Π_{Gen} is the generator loss and $\Pi_{Perceptualloss}$ is the perceptual loss. Π_{G}^{Ra} is called the adversarial loss which is the loss of a relativistic generator, $L1$ is the content loss and Π_{Dis}^{Ra} is the discriminator loss. μ and η represent the coefficients to offset the losses.

Perceptual loss

Perceptual loss works to improve the texture and picture accuracy of the generated images [42]. Euclidean distance is used to compare the feature maps of the original image $\hat{v}^{k+1}$ and the generated image v^{k+1}. According to the definition of [42], the feature map was extracted before using the generator network's final activation function. The extraction of feature maps after activation function caused the model to be inconsistent, directly impacting the model output.

When recapturing HR from LR, it provides close supervision between feature maps. The fact that scalograms are not sufficiently HR is well understood, and this aspect boosts model re-generation dramatically. Mapping feature α_{ij} is gotten after j^{th}-convolution and before the max-pooling layer. The formality is measured as the distance between the function representations of the super-resolution image $G_k\left(v^k\right)$ and the real image $\hat{v}^{k+1}$. Formal calculation between feature maps is given in (2).

$$\begin{aligned}\Pi_{Perceptualloss} =& \sum_{x=1}^{W_{ij}}\sum_{y=1}^{H_{ij}}\left(\alpha_{ij}(\hat{v}^{k+1})_{xy}\right. \\ &\left.- \alpha_{ij}(G_k(v^k))_{xy}\right)^2\end{aligned} \tag{2}$$

Rather than encouraging the pixels of the output image v^{k+1} to exactly match the pixels of the target image $\hat{v}^{k+1}$, perceptual loss encourages them to have similar feature representations as computed by the loss network.

Content loss

By manipulating the HR image v^{k+1} to be close to the ground truth $\hat{v}^{k+1}$, the network improves the accuracy of pixel-level by calculating the $L1$-norm distance between both the ground truth and the recovered image. (3) calculates the $L1$-norm distance between the SR image $\left(G_k(v^k)\right)_{xy}$ and the ground truth $(\hat{v}^{k+1})_{xy}$ are given in (3).

$$Ł1 = \sum_{x}^{W}\sum_{y}^{H}\left\|G_k(v^k)_{xy} - (\hat{v}^{k+1})_{xy})\right\|_1 \tag{3}$$

Relativistic loss

The majority of the preliminary research focused on standard GAN. Meanwhile, we employ a rational discriminative loss in our SR network, ensuring that HR photos are not stylized or unrealistic. In (4), the classification of the images uses the standard discriminator D_{is} in GAN.

$$\begin{aligned}D_{is} &= \sigma\left(f_d(\hat{v}^{k+1})\right) \to 1 \\ D_{is} &= \sigma\left(f_d(v^k)\right) \to 0\end{aligned} \tag{4}$$

Equation (4) reflects the regular GAN's operation. D_{is} is the discriminator's output to classify whether the images are real or artificial. The vector feature discriminator is represented as $f_d(.)$. Additionally, the word "σ" stands for the sigmoid function. Adversarial loss is a binary classifier that differentiates between real data and generated data predicted by the generative network. We use the relativistic GAN [36] to distinguish between the real $\hat{v}^{k+1}$ and created data $G_k(v^k)$ with the distance computed in (5).

$$D_{Ra}\left(\hat{v}^{k+1}, G_k(v^k)\right) \tag{5}$$

RGAN produces images with sharp edges when used in a relativistic model and provides more graphic and detail information than a typical GAN as presented in (6).

$$\begin{aligned}&D_{Ra}(Real, Fake) = C(Real) - E\left(C(Fake)\right) \to 1 \\ &\text{how realistic an image is compared to a fake one.} \\ &D_{Ra}(Fake, Real) = C(Fake) - E\left(C(Real)\right) \to 0 \\ &\text{how fake an image is compared to a real one.}\end{aligned} \tag{6}$$

Here, $E(.)$ is the average of all real or fake data in the sample. This slight modification makes the model more efficient than the standard discriminator network. The discriminator network loss is given in (7).

$$\begin{aligned}\Pi_{Dis}^{Ra} = &- E_{\hat{v}^{k+1}}[log\big(D_{Ra}(\hat{v}^{k+1}, G_k(v^k))\big)] \\ &- E_{G_k(v^k)}[log\big(1 - D_{Ra}(G_k(v^k), (\hat{v}^{k+1})\big)]\end{aligned} \tag{7}$$

Despite this, (8) illustrates the adversarial loss for the RGAN.

$$\begin{aligned}\Pi_{G}^{Ra} = &- E_{\hat{v}^{k+1}}[log\big(1 - D_{Ra}(\hat{v}^{k+1}, G_k(v^k))\big)] \\ &- E_{G_k(v^k)}[log\big(D_{Ra}(G_k(v^k), (\hat{v}^{k+1})\big)]\end{aligned} \tag{8}$$

The network is concurrently trained for both actual image $\hat{v}^{k+1}$ and created image $G_k(v^k)$ to minimize the failure of the discriminator and generator networks. When the discriminator reaches the optimal value, the gradient gets close to zero which provides little feedback to the generator, thereby slowing or completely stopping the learning. At this level, custom GAN does not learn how to create more realistic images. In comparison, RGAN study both images and gradient are dependent on both terms, i.e., $\hat{v}^{k+1}$ and $G_k(v^k)$.

Modified wavelet convolutional capsule network (MWCCN)

Presently, research literature have shown that CNN have generated excellent results in the extraction of features for classification problems. Conventional CNNs use scalar neurons to express the likelihood of distinguishing features being present, which severely restricts their performance. Figure 2 illustrates a fine-tuned VGG-19 model with discrete wavelet transform (DWT) pooling used for the feature extraction of scalograms. Traditional VGG-19 has 16 convolution layers and 3 fully connected layers. To this, we executed few modification using the pre-trained weight while keeping the first block and replaced subsequent blocks having max-pooling with the discrete wavelet pooling.

As low-level features such as curves, color, edges and texture are extracted from the first block, thus high-level properties are extracted as the network goes deeper. However, the fundamental goal is to replaced max-pooling with DWT pooling to reduce the loss of spatial details. In this study, after the feature extraction stage, we discarded the 3 fully connected layers in the pre-trained VGG 19, thus having the last block layer with $14 \times 14 \times 512$ feature vector as output. For dimensionality match, this last block layer is connected to the primary capsule layer of the capsule network using a $14 \times 14\times$ convolutional layer with kernel size of 256 and stride of 3 represented as f_1 in Fig. 2.

Loss of spatial information is one of the causes for CNN's low classification efficiency. To properly identify and categorize bearing defects, we suggested a Modified Wavelet Convolutional Capsule Network (MWCCN). Capsule network for categorization problems was first proposed by the authors in [43]. Unlike standard CNNs, a capsule network consists of capsules of vectorial entities. A capsule is a group of neurons that are arranged in a vectorial pattern [44]. The starting parameters of a capsule represent a specific class of entity, and also the length of the capsule indicates the chances that the entity exists. Capsule networks outperform regular CNN in obtaining intrinsic and differentiating features of entities [43]–[45].

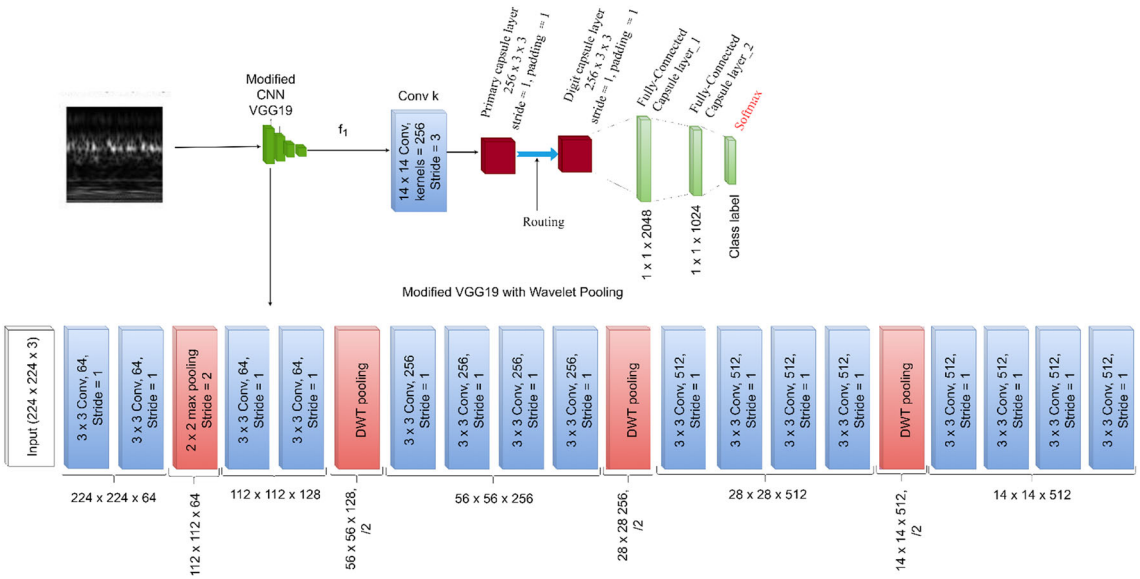

Fig. 2 Detailed structure of our proposed modified wavelet convolutional capsule network (MWCCN)

To attain excellent fault detection and classification efficiency, we adapt the original capsule model and fine-tuned the network by including a pre-trained VGG-19 framework to extract features in order to develop a deep convolutional capsule framework. Figure 2 shows a concise representation of the proposed wavelet convolutional capsule framework. Low-level attributes of the scalograms such as curves, color, edges, and texture are obtained from the few convolutional layers at the beginning stage of the network while the high-level attributes are obtained as the convolutional layers grow deep. To improve the classification performance and maintain the integrity of high-level features, we implemented a pooling operation called the discrete wavelet transform (DWT) pooling to achieve down-sampling. This minimizes the loss of spatial information and allows for dimension reduction. After extracting the features from the CNN framework, the features are passed to another convolutional layer to achieve a match in dimension before reshaping the features into primary capsules. Routing by agreement is applied to map the features between the primary and digit capsules. As shown in equation (9), the total input in the capsule layers consists of sum total of the weights of all predictions obtained from the capsules within the capsule network.

$$C_j = \sum_i a_{ij} \, . \, U_{j/i} \tag{9}$$

Where C_j depicts the entire input to capsule j. a_{ij} is the coupling coefficient which indicates the level to which capsule i ignites capsule j. $U_{j/i}$ is the prediction of capsule j from capsule i as illustrated in (10).

$$U_{j/i} = W_{ij} \, . \, U_i \tag{10}$$

W_{ij} represents the network weight mapping capsule i to j whereas U_i represents the output of capsule i. A routing by agreement algorithm decides the coefficient between the primary and digital capsules summing to 1 [43]. This routing approach takes into account both the length and representation parameters of the capsule and when igniting another capsule whereas in conventional CNN a framework depends on the evaluated probability. In a nutshell, capsule networks have a better dependency and capability of abstracting distinct inherent features. It's worth noting that the capsule's length is utilized to assess the possibility of the existence of an entity. For a perfect probability prediction, a non-linear activation function called squashing function is applied, where capsules with short vectors are marked as low probability and that of long vectors are marked as high probability, while retaining a fixed orientation. (11) gives the squashing function formula.

$$U_j = \frac{||C_j||^2}{1 + ||C_j||^2} \cdot \frac{C_j}{||C_j||} \tag{11}$$

The high-level entity abstraction is further passed into the 2 fully connected layers, and then a Softmax classifier is used for the classification task. For a successfully training the capsule network for classification tasks, margin loss [43] is applied. Equation (12) defines the margin loss, L_k for class$k$.

$$\begin{aligned} L_k =& T_k \, . \, max(0, m^+ - U_k)^2 \\ &+ \eta(1 - T_k) \, . \, max(0, U_k - m^-)^2 \end{aligned} \tag{12}$$

In the Softmax layer, U_k is represented as the output of the capsule. If the training sample is an instance of class k, then T_k is set to 1 , else, T_k is set to 0. m+ is set to 0.9 and m- is set to 0.1 which represent the lower and upper bounds for the probability of a training data becoming or not becoming an instance of class k respectively.n is the weight regularizer which is by default set to 0.5. The total loss of the capsule network is the summation of all digit capsule losses.

The Proposed Modified Wavelet Convolutional capsule Network with MESRGAN+ (MWCCN-MESRGAN+)

Our proposed MWCCN-MESRGAN+ is an integrated super resolution GAN and wavelet convolutional capsule network for diagnosing rolling bearing fault as presented in Fig. 3. The proposed architecture consists of the super resolution part which handles the image enhancement by reconstructing high-resolution images from low-resolution image counterpart as the first stage while the second part is the wavelet convolutional capsule network which extracts and learns high dimensional feature vectors from the super resolution imagery generated by the super resolution network for fault diagnosis and classification. We adopted some evaluation metrics such as receiver operating characteristic (ROC), accuracy (ACC), sensitivity, specificity (SPE), and precision (PRE). Details of the dataset utilized in the paper are described in Sect. Experiments .

Constructing the time-frequency scalograms

Convolutional neural network processes images in either 2D grayscale or RGB. To this end, the raw fault signals are converted to time-frequency scalograms of 2D images with abundant fault features using CWT. This paper adopts cmor3-3 wavelet for CWT due to its excellent ability to analyze time-frequency. The raw signals are collected at each sample point of 1024 in time series and CWT is also executed at the same time series.

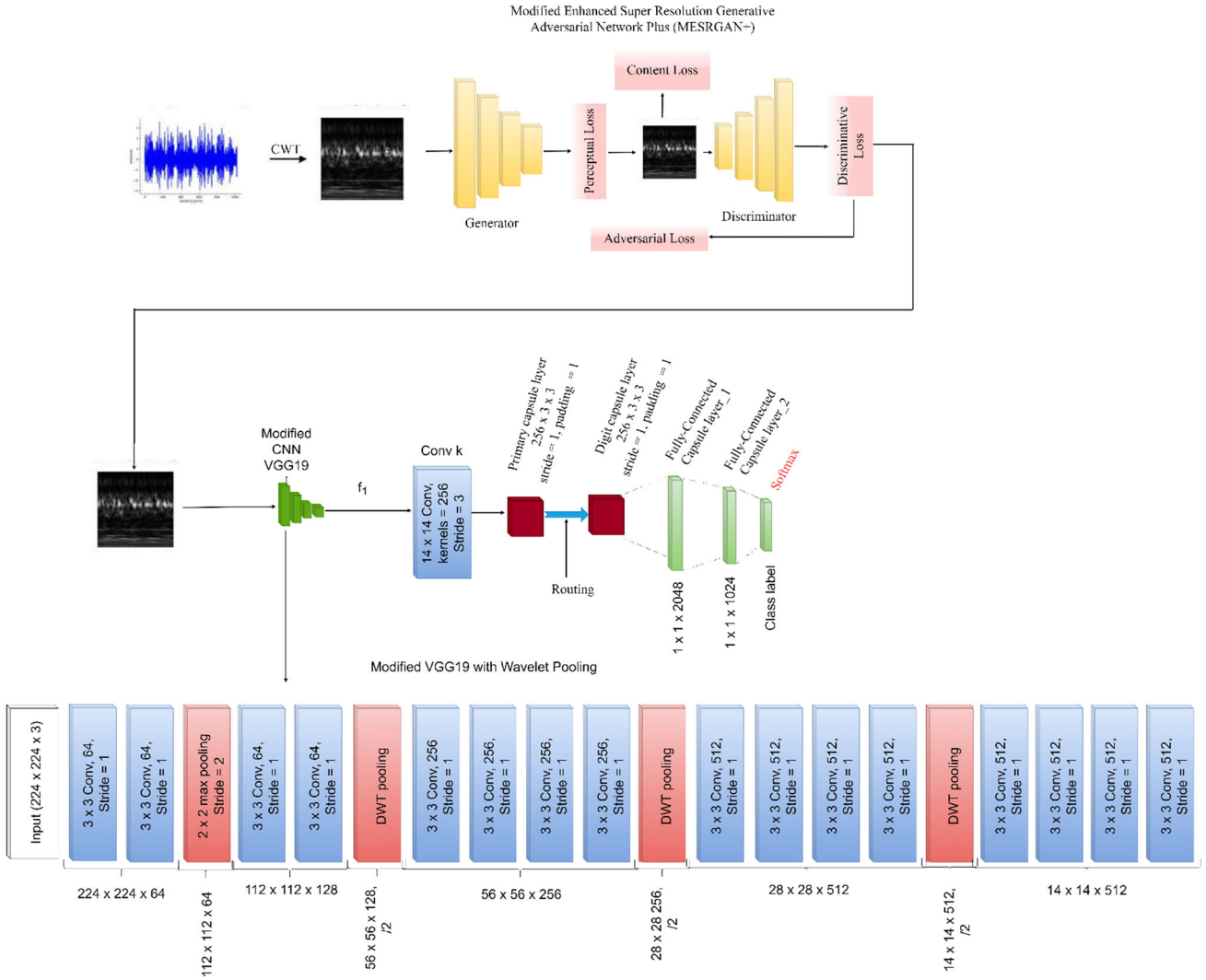

Fig. 3 The proposed modified wavelet convolutional capsule network with modified enhanced super resolution generative adversarial network plus (MWCNN-MESRGAN+)

Experiments

Experimental Setup and Details

We built our proposed model using the famous rolling bearing dataset of CWRU [38]. The vibration signal of the bearing motor is collected using acceleration sensor. The bearing dataset collection is structured as normal, drive end fault with sampling frequency of 12 kHz and 48 kHz and fan end fault with sampling frequency of 12 kHz. The drive end defect with 12 kHz frequency is utilized in this paper. This data was measured under different load conditions (0, 1, 2, and 3 hp) as a single point fault due to electro-discharge machine. It is partitioned into ball fault (BF), inner fault (IF), outer fault (OF) with different fault diameters (0.007, 0.0014, and 0.0021 inches). Besides, the outer race of the dataset consists of 3 damage points which are 3 o'clock, 6 o'clock, and 12 o'clock. The dataset consists of 10 health conditions for which the 9 conditions (BF-7, BF-14, BF-21, IF-7, IF-14, IF-21, OF-7, OF-14, OF-21) belong to the fault label and 1 condition belongs to the normal label (N).

This paper adopts the damage point of 6 o'clock and the load condition of 0 hp to build the data split for our proposed model. Since there are 10 conditions with 1024 signal sampling points, we randomly select 400 samples to build the training set and 200 samples to build the testing set for each condition which will eventually amount to a total of 4000 training set and 2000 testing set. Additionally, a validation set is constructed from 25% of the training set which bring the training set to 3000 samples and 1000 samples for the validation set. Since the proposed model requires that the input data must be a 2D image with three channels (RGB), we adopted CWT to convert the raw signal into time-frequency scalograms (2D images) and reshaped the input dimension

to 224 × 224 × 3. It is important to mention that Adam optimizer and the learning rate of 0.0002 are used with a batch size of 16 and 30 epochs during the training of our model and the number of class label in the dataset correspond to 10. We implement our proposed model on Keras framework with Tensorflow as backend using NVIDIA GTX 1080 GPU work station.

Evaluation

This section presents the results of our study in two parts; the first part presents the evaluation of our proposed modified super resolution GAN plus network in terms of peak signal to noise ratio(PSNR) and Perceptual Index (PI). The second part involves evaluating the fault diagnosis performance of our proposed wavelet convolutional capsule network (MWCCN). We made few comparison with state-of-the-art models including some selected pre-trained models. The evaluation criterion adopted as the metric to evaluate the diagnosis performance of our proposed MWCCN is as follows: accuracy (ACC), precision (PRE), sensitivity (SEN), specificity (SPE), area under curve (AUC) and F1-Score.

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{15}$$

$$Specificity = \frac{TN}{TN + FP} \tag{16}$$

where TP, FP, and FN indicates the outcomes of true positive, false positive, and false negative, respectively.

MSRGAN+ Evaluation

Table 1 illustrations the efficacy of our proposed MESRGAN+ model in terms of PSNR and PI as well as the transition of SRGAN to MESRGAN+ in comparison with other well-known methods. The MESRGAN+ generates more suitable images, eliminates unimportant details and artifacts, and enhances extracting feature visibility. The time-frequency scalograms are fed into the proposed MESRGAN+ for low quality enhancement and achieving high-resolution (HR) by up-scaling with a factor of 4.The HR images created by the MESRGAN+ preserve abundant fault features while discriminating against distracting backgrounds.

The results of the CWT conversion of raw signals to scalograms are presented in Fig. 4 (load condition = 0 hp). According to Fig. 4, the fault types that correspond to the raw time domain signals is difficult to distinguish. However, CWT makes it easy to distinguish the differences between the time-frequency scalogram of individual fault category which makes it suitable for our proposed MESRGAN+ to extract abundant features for image regeneration.More so, Fig. 5 shows the performance of our proposed super-resolution, MESRGAN+ and other state-of-the-art models which are SRGAN, ESRGAN and ESRGAN+. For fair comparison, we employed their source codes available online with the same CWRU dataset. One of the aims of this research is to check the PSNR and perceptual index (PI) of the super-resolution models in which our model gives the bet results in both cases. MESRGAN+ produces more appropriate images, removes artifacts, and improves extracting features clarity by extending the convolutional layer of the generative structure of the residual block and removing batch normalization.

Fault diagnosis evaluation of MWCCN

Results

The accuracy and loss curves of our proposed MWCCN-MESRGAN+ are presented in Figs. 6 and 7. We observed that after 10 epochs, the training and validation curves started converging with smooth stability depicting efficacy of the model. More so, the diagnostic accuracy of the model is achieved using the trained model to classify the testing dataset. To ensure the stability of our proposed model, the work was repeated for 7 times under same condition and the

Table 1 Comparison of the structural configuration of SRGAN, ESRGAN, ESRGAN+ and our proposed MESRGAN+ including their reported PSNR and Perceptual Index using data-class A

Parameter	SRGAN	ESRGAN	ESRGAN+	MESRGAN+
Residual block of the generator	Conv (3,64,1) Batch norm ReLU Conv(3,64,1) Batch norm	Conv (3,64,1) ReLU Conv (3,64,1)	Conv (3,64,1) ReLU Conv (3,64,1)	Conv (3,64,1) ReLU Conv (1,64,1) ReLU Conv (3,64,1) ReLU Conv (3,64,1)
Input size	LR	LR	LR	LR
PSNR	19.34 dB	19.17dB	18.56 dB	18.36 dB
Perceptual index	2.87	2.61	2.24	2.18

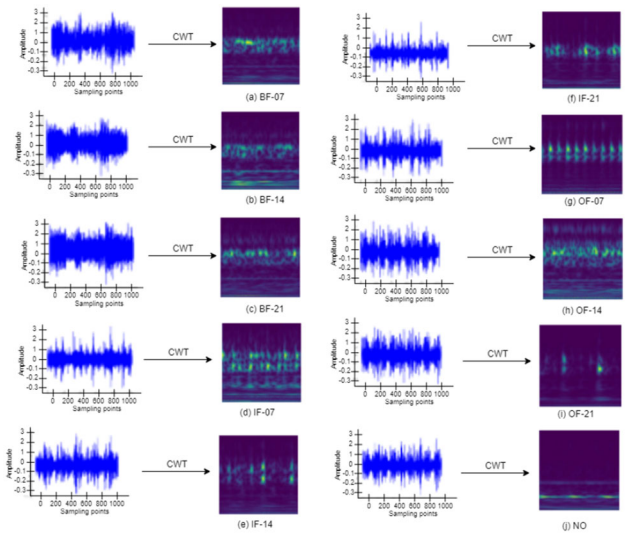

Fig. 4 The construction of time-frequency scalograms under 0 hp load condition using CWT

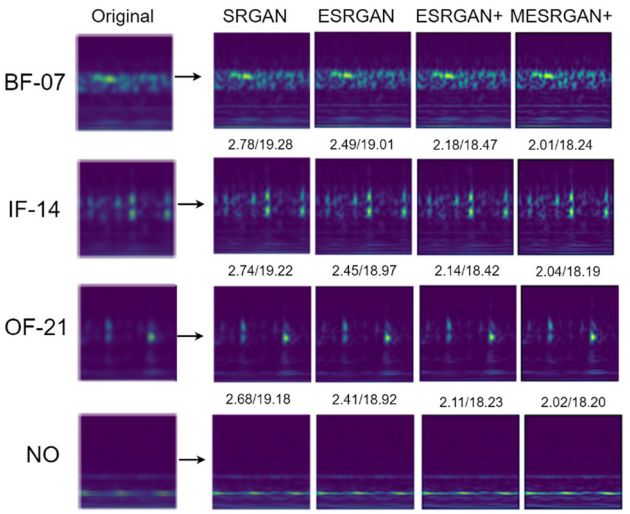

Fig. 5 Super-resolution results of our modified version of ESRGAN+ in comparison with state-of-the-art models. The PI score is shown on the left while the PSNR score is shown on the right

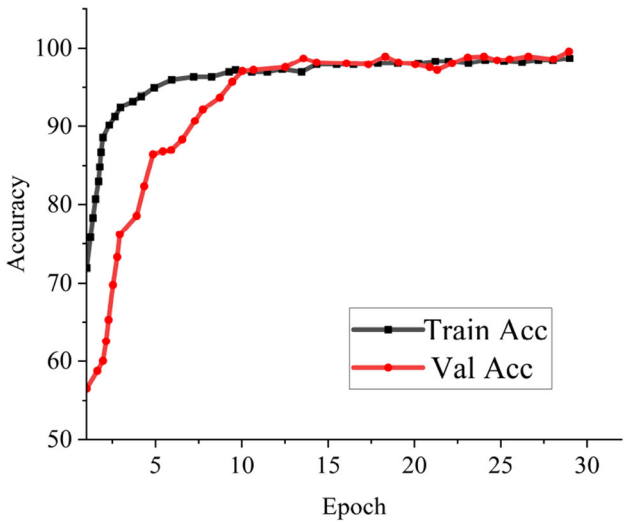

Fig. 6 The accuracy curve of the proposed MWCCN-MESRGAN+ using the data-class A category

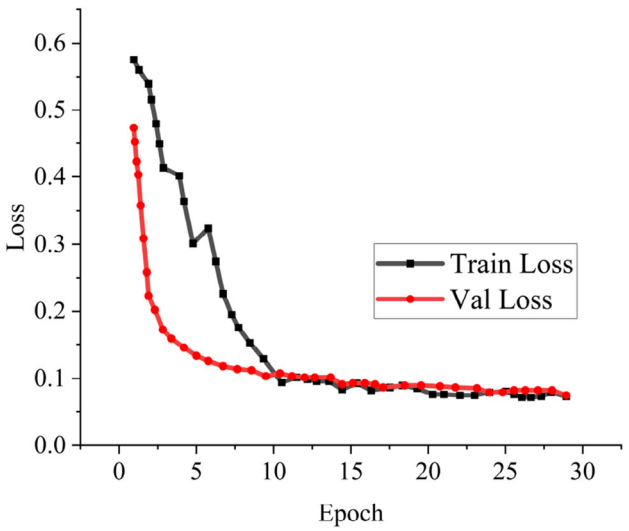

Fig. 7 The loss curve of the proposed MWCCN-MESRGAN+ using the data-class A category

accuracy, sensitivity, specificity, and precision presented in Table 2. As presented in Table 2, the proposed model obtained 99.92% accuracy, 99.78% sensitivity, 99.69% specificity, and 100% precision for the first category of sub-data class under the load condition of 0 hp indicating the model's robustness in fault diagnosis.

Ablation study

Model tweaking

For the purpose of understanding the influence of super resolution approach and discrete wavelet pooling on image quality enhancement and the performance of the proposed MWCCN-MESRGAN+ model, we conducted some experiment by making some structural adjustment to our proposed MWCCN-MESRGAN+ model as shown in Table 3. The four different models are trained on the first category of sub-data class called data-class A and the training epochs is set to 30.

All training parameters are kept the same and the training time for each models to complete one epoch is recorded and the results are given in Table 4. We observed that the time required for the proposed MWCCN-MESRGAN+ model to train for one epoch is 9s, 16s, and 20s longer than the other models as shown in Fig. 8. This is a clear indication that more parameters more training time. More to the point, MWCCN-

Table 2 The result of our proposed MWCCN-MESRGAN+ on the data-class A

Data-class	ACC (%)	SEN (%)	SPE (%)	PRE (%)	Computational time (s)
A	99.92	99.78	99.69	100	2,790

Table 3 The result of the experiments conducted by adjusting the structural parameters of our proposed model on data-class A

Model	ACC (%)	SEN (%)	SPE (%)	PRE (%)
MWCCN-MESRGAN+ (Proposed)	99.92	99.78	99.69	100
MWCCN (max-pool)-MESRGAN+	98.04	98.11	97.98	98.32
MWCCN (DWT)	97.71	98.34	98.85	98.92
MWCCN (max-pool)	96.58	97.78	97.96	98.43

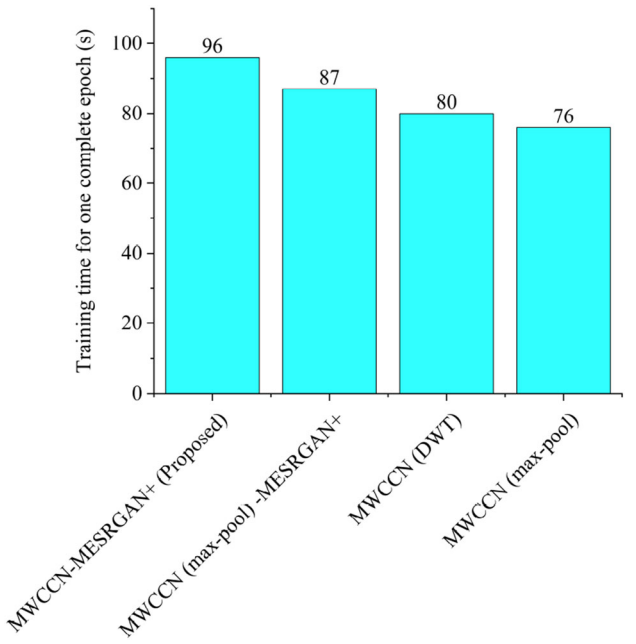

Fig. 8 The average training time for the different models constructed from adjusting the proposed MWCCN-MESRGAN+. This computational time is recorded for one complete epoch using the data-class A category

MESRGAN+ converges fast with higher accuracy and lower loss despite the long training time.

Investigating dataset generalization. NC stands for normal condition. LC stands for load condition. HC stands for Health condition

Most deep learning models utilize max-pooling layers to perform down-sampling operation to reduce the dimensionality of the feature vector but this process usually lead to loss in spatial features although the computational time is reduced. We know that capsule network has a longer training time due to the size of its feature dimension, however, it maintains the integrity of its high-level feature without loss of spatial feature which gives capsule-based network the competitive advantage over traditional convolutional neural networks to learn specific features from the dataset resulting to higher accuracy and fast convergence as a trade-off between computational time and accuracy. It is worth mentioning that our proposed MWCCN-MESRGAN+ model achieves nearly 100% diagnosis accuracy without overfitting on small dataset.

Investigation of dataset generalization

To further investigate the generalization of our proposed MWCCN-MESRGAN+ model as presented in Table 4, we constructed several categories of sub-data class under different load conditions as follows;

Data-class B 400 samples are selected randomly under the load condition of 1hp for individual health condition as the training set, and 200 samples are selected randomly under the same load condition for the testing set. For the 10 conditions, the total training samples and testing samples are 4000 and 2000 respectively.

Data-class C 400 samples are selected randomly under the load condition of 2hp for individual health condition as the training set, and 200 samples are selected randomly under the same load condition for the testing set. For the 10 conditions, the total training samples and testing samples are 4000 and 2000 respectively.

Data-class D 400 samples are selected randomly under the load condition of 3hp for individual health condition as the training set, and 200 samples are selected randomly under the same load condition for the testing set. For the 10 conditions, the total training samples and testing samples are 4000 and 2000 respectively.

Data-class E 200 samples are selected randomly under the load condition of (0 and 1 hp) for individual health condition as the training set, and 100 samples are selected randomly under the same load condition for the testing set. For the 10 conditions, the total training samples and testing samples are 4000 and 2000 respectively.

Data-class F 150 samples are selected randomly under the load condition of (0, 1, and 2 hp) for individual health condition as the training set, and 80 samples are selected randomly under the same load condition for the testing set. For the 10 conditions, the total training samples and testing samples are 4500 and 2400 respectively.

Data-class G 100 samples are selected randomly under the load condition of (0, 1, 2, and 3 hp) for individual health

Table 4 Investigating dataset generalization. NC stands for normal condition. LC stands for load condition. HC stands for Health condition

	Data-class B		Data-class C		Data-class D		Data-class E				Data-class F						Data-class G						Data-class H					
	LC: 1hp		LC: 2hp		LC: 3hp		LC: 0hp		LC: 1hp		LC: 0hp		LC: 1hp		LC: 2hp		LC: 0hp		LC: 1hp		LC: 2hp		LC: 3hp		LC: 0hp		LC: 1hp	LC: 2hp
HC	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts
NC	400	200	400	200	400	200	200	100	200	100	150	80	150	80	150	80	100	50	100	50	100	50	100	50	100	50	100	200
BF-7	400	200	400	200	400	200	200	100	200	100	150	80	150	80	150	80	100	50	100	50	100	50	100	50	100	50	100	200
BF-14	400	200	400	200	400	200	200	100	200	100	150	80	150	80	150	80	100	50	100	50	100	50	100	50	100	50	100	200
BF-21	400	200	400	200	400	200	200	100	200	100	150	80	150	80	150	80	100	50	100	50	100	50	100	50	100	50	100	200
IF-7	400	200	400	200	400	200	200	100	200	100	150	80	150	80	150	80	100	50	100	50	100	50	100	50	100	50	100	200
IF-14	400	200	400	200	400	200	200	100	200	100	150	80	150	80	150	80	100	50	100	50	100	50	100	50	100	50	100	200
IF-21	400	200	400	200	400	200	200	100	200	100	150	80	150	80	150	80	100	50	100	50	100	50	100	50	100	50	100	200
OF-7	400	200	400	200	400	200	200	100	200	100	150	80	150	80	150	80	100	50	100	50	100	50	100	50	100	50	100	200
OF-14	400	200	400	200	400	200	200	100	200	100	150	80	150	80	150	80	100	50	100	50	100	50	100	50	100	50	100	200
OF-21	400	200	400	200	400	200	200	100	200	100	150	80	150	80	150	80	100	50	100	50	100	50	100	50	100	50	100	200
Total	4,000	2,000	4,000	2,000	4,000	2,000	2,000	1,000	2000	1000	1500	800	1500	800	1500	800	1000	500	1000	500	1000	500	1000	500	1000	1000	1000	2000

Table 5 The results of MWCCN-MESRGAN+ model under different load conditions for the rolling bearing dataset

Data-class	ACC (%)	SEN (%)	SPE (%)	PRE (%)
B	99.92	99.78	99.69	100
C	99.04	99.11	99.98	99.98
D	99.71	99.34	99.85	100
E	99.85	99.78	99.96	99.97
F	99.58	99.86	99.97	100
G	99.67	99.89	99.89	99.99
H	99.91	99.85	99.95	100

condition as the training set, and 50 samples are selected randomly under the same load condition for the testing set. For the 10 conditions, the total training samples and testing samples are 4000 and 2000 respectively.

Data-class H 100 samples are selected randomly under the load condition of (0 and 1 hp) for individual health condition as the training set, and 200 samples are selected randomly under the load condition of 2hp for the testing set. For the 10 conditions, the total training samples and testing samples are 4000 and 2000 respectively.

Additionally, 25% split of the training set for each data-class category above is utilized to build the validation dataset during training. the proposed model is used to train the data-class B - H. The training parameters and conditions are kept the same as the previous data-class A. The proposed MWCCN-MESRGAN+ is trained for 7 times repeatedly and the experimental outcome is presented in Table 5. Damage point of 6 o'clock is utilized for all data-class categories. The results indicates that our proposed model relatively achieved nearly 100% across the different categories of data-class.

It is worth mentioning that The proposed MWCCN-MESRGAN+ achieves excellent diagnosis accuracy fault data under different load conditions. The data-class H is a special construction to test the validation of our proposed MWCCN-MESRGAN+ by introducing a new load condition from 2 hp as the model is trained on the load condition (0, 1 hp). However, the model still classified the bearing faults which indicates that our proposed MWCCN-MESRGAN+ model is robust and generalizes very well.

Comparison with other MWCCN (Inception V3, EfficientNet) Models

Our proposed MWCCN in this paper is constructed using VGG-19 as the base model. To examine the efficacy of MWCCN (VGG-19), We compared the model with other MWCCN(Inception V3 and EfficientNet) models. We only fine-tuned the last layer of the pre-trained models by replac-

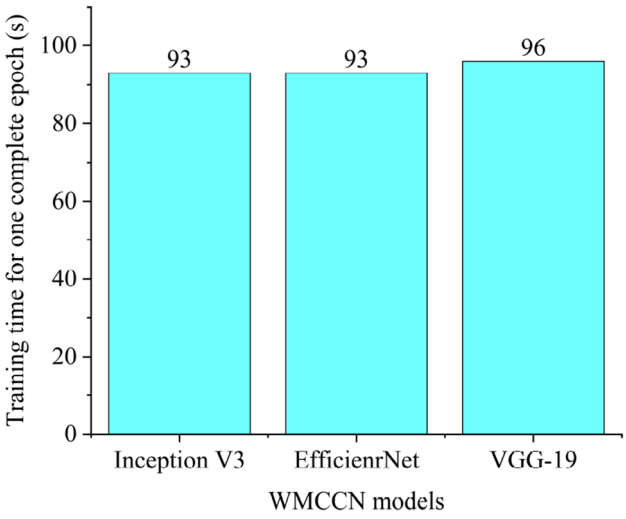

Fig. 9 The average training time for the different MWCCN models. On one hand, MWCCN is modified using Inception V3 as the base model. on the other hand, EfficientNet is used as the base model. We compared these two modifications with our proposed model with VGG-19 as the base model

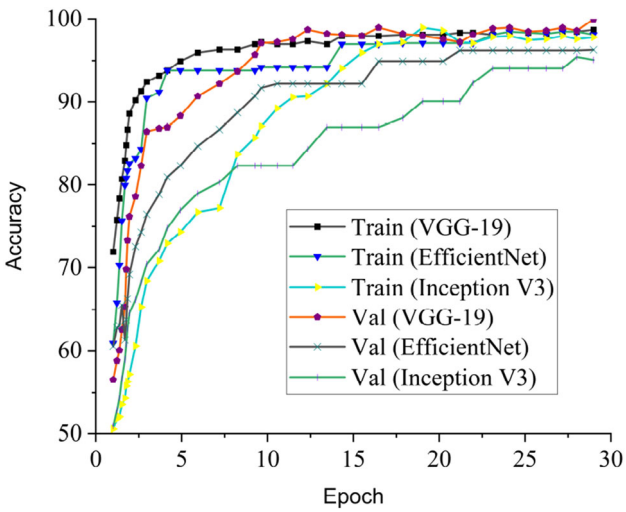

Fig. 10 The accuracy curves of the various MWCCN models on data-class A category

ing the neurons in the Softmax layer to correspond to the number of class label in our dataset which in our case is 10 neurons. The different versions of the constructed MWCCN models including our proposed model are trained on the first sub-dataset category, data-class A. The computational time for training one epoch by the three models are presented in Fig. 9.

Figure 10 shows the accuracy graph for both the training and validation dataset for the individual MWCCN model. We observed that MWCCN (Inception V3), MWCCN (EfficientNet) and MWCCN (VGG-19) requires approximately the same training time to complete one epoch but MWCCN (VGG-19) converges faster. In a bid to further ascertain the excellent performance of our proposed MWCCN (VGG-19), we conducted another experiment on all the categories of the sub-dataset (Data-class A-H). Table 6 depicts that MWCCN (VGG-19) achieves excellent diagnosis performance across the sub-datasets depicting that MWCCN (VGG-19) generalizes better than MWCCN (Inception V3) and MWCCN (EfficientNet).

Comparison with other fault diagnosis methods

In the course of our work, we reviewed several literature related to fault diagnosis based on artificial intelligence and presented some comparison. Some literature reported few performance indicators to support their claims as seen in Table 7. More to the point, our proposed model achieves better performance with more indictors reported compared to the other fault diagnosis methods cited from literature. the proposed MWCCN-MESRGAN+ model is compared with other fault diagnosis methods.

The authors in [44] proposed a wavelet based multi-fractal feature learning approach with SVM classifier. The authors in [45] adopted the method of ELM to diagnose bearing faults. An interesting work was proposed in [46] which is based on SVM and EEMD for fault diagnosis. The authors in [47] integrated wavelet into auto-encoder learning and combined the framework with ELM to diagnose faults. The authors in [48] suggested a solution to the problem of hierarchical recognition in machine by using DBN to diagnose fault. Quite

Table 6 The comparison of MWCCN models under different load condition

MWCCN model	Data-class A ACC (%)	Data-class B ACC (%)	Data-class C ACC (%)	Data-class D ACC (%)	Data-class E ACC (%)	Data-class F ACC (%)	Data-class G ACC (%)	Data-class H ACC (%)
Inception V3	99.76	99.58	98.97	99.63	99.61	99.32	99.49	99.76
EfficientNetl	99.84	99.47	98.82	99.69	99.74	99.46	99.51	99.87
VGG-19	99.92	99.65	99.04	99.71	99.85	99.58	99.67	99.91

Table 7 Comparison results of our proposed model with other fault diagnosis methods

Method	Data-class A ACC (%)	Data-class B ACC (%)	Data-class C ACC (%)	Data-class D ACC (%)	Data-class E ACC (%)	Data-class F ACC (%)	Data-class G ACC (%)	Data-class H ACC (%)
Wavelet features + SVM [44]	88.90	–	–	–	–	–	–	–
ELM [45]	97.50	–	–	–	–	–	–	–
EEMD + SVM [46]	97.64	99.12	99.64	97.91	–	–	–	–
DWAE + ELM [47]	95.20	–	–	–	–	–	–	–
DBN [48]	99.03	–	–	–	–	–	–	–
SSAE + DNN[49]	–	–	–	–	99.10	–	–	–
CNN + RF [50]	–	–	–	–	99.73	99.08	–	–
Proposed method	99.92	99.65	99.04	99.71	99.85	99.58	99.67	99.91

an interesting work of DNN with SSAE was suggested by the authors in [49] for the diagnosis of fault.

The authors in [26] utilized 2D and CNN for the diagnosis of bearing faults. Random forest learning with CNN was suggested by the authors in [50] to diagnose faults. We observed that our proposed method is superior to conventional machine learning approach and other fault diagnosis model. Compared with the methods suggested by the authors in [44] and [45], our proposed MWCCN-MESRGAN+ model showed a significant improvement by a large margin. Compared with the other methods including the deep learning model, our proposed model significantly outweighs all of the models. Compared to the method in [50] under the data-class F, the accuracy of our method increased by 0.5% showing that our model generalizes very well.

The experimental results show that our proposed architecture outweighs other fault diagnosis models and some selected deep learning models. For fairness, we selected some deep learning models and implemented them based on their source code using the same data-class A. From the experimental analysis of our comparative report as presented in Table 8. MobileNet V2 achieves the least sensitivity score of 93.6% whereas ResNet50 obtains the least specificity score of 92.5% as depicted in Table 8. From all indications, our proposed model outweighs all the pre-trained models with a high sensitivity score of 99.78% and 99.69% specificity score. Another important metric is the Receiver Operating Characteristics (ROC) curve. The ROC curve measures the overall accuracy in terms of AUC as shown in Fig. 11. Fig. 11 shows that our model demonstrates a satisfactory balance between sensitivity and specificity by minimizing the error rate of the false positive and maximizing the true positive rate.

More so, the accuracy performance of the pre-trained models is reported in comparison with our proposed model as presented in Table 8. Our model performs better than the pre-trained models, achieving a high accuracy of 99.92%. We also show that our proposed model convergences smoothly and steadily with a moderate reduction in loss.

Conclusion

In this work, we proposed a GAN-based super-resolution with wavelet convolutional capsule framework for fault diagnosis and classification with the aim of handling the challenge of low-quality characteristics of scalograms obtained from raw fault signals by using continuous wavelet transform (CWT). In summary of the contribution of this work, CWT is used to convert the raw fault signals into 2D time-frequency scalograms compatible for 2D CNN operation. The 2D time-frequency images (scalograms) are reshaped to $224 \times 224 \times 3$

Table 8 Comparison results of our proposed model with selected transfer learning models on data-class A

Model	ACC (%)	SEN (%)	SPE (%)	PRE (%)
MWCCN-MESRGAN+ (Proposed)	99.92	99.78	99.69	100
MobileNet-V2	94.71	93.60	94.28	96.72
ResNet-50	95.02	94.72	92.50	95.82
DenseNet-121	94.98	95.68	93.87	96.43
Xception	96.56	96.82	95.62	97.66

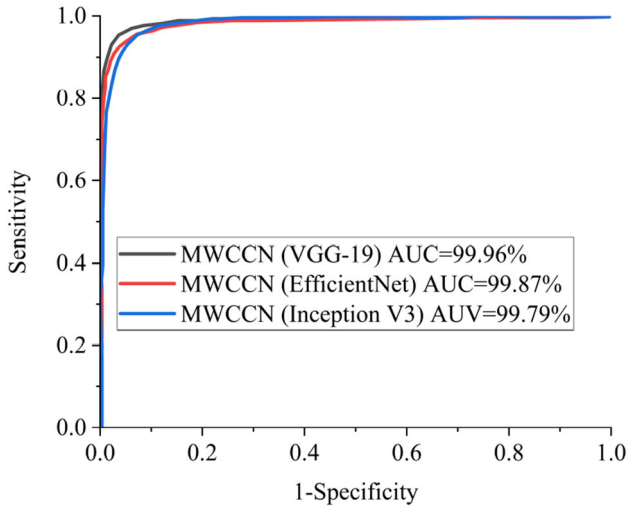

Fig. 11 The ROC curves of the various MWCCN models on data-class A category

RGB format as input to the GAN-based super resolution network for quality enhancement.

The reconstructed high-resolution images become the new input images to the wavelet convolutional capsule network for fault diagnosis and classification. More so, the proposed MWCCN-MESRGAN+ model is validated with the famous rolling bearing fault dataset from CWRU achieving 99% accuracy, 99% specificity, 99% sensitivity and 100% precision which outweighs the other fault diagnosis methods including some deep learning models. We carried out ablation study to evaluate the generalization performance of our proposed model and by a well-observed margin, the results demonstrate that our proposed MWCCN-MESRGAN+ model achieved excellent fault diagnosis performance.

Even though this study has a high level of accuracy in classifying fault, it does have certain drawbacks. This suggested strategy, which has high classification accuracy in CWRS dataset, might not obtain exactly the same classification accuracy in imbalanced fault dataset. The reason is because the class labels may consist of imbalanced data samples owing to differences in labeling. To solve this challenge, AI model should be trained utilizing imbalanced class label data acquired at various times and locations. Aside the diversity of data, the allocation of the data classes is also significant. The disparity in class sizes has a detrimental impact on training. The accuracy of classification is also affected by the different data augmentation strategies employed to correct the imbalance. In light of this constraint, study will be conducted in our future work employing a wider range of imbalance class data and possibly employing various optimization strategies that are more efficient in terms of computation time.

Author Contributions HNM: Conceptualization, Methodology, Resources, Data curation, Formal analysis, Writing - original draft. JPL: Project administration, Supervision. GUN: Investigation, Software, Validation, Visualization, Writing - review and editing. MAH, AO: Validation, Visualization, Writing - review and editing.SN, JJ: Software, Validation, Writing—review and editing.

Data Availability The datasets utilized in this work are publicly available in the following links (https://engineering.case.edu/bearingdatacenter/download-data-file) (https://csegroups.case.edu/bearingdatacenter/home)

Declarations

Conflict of interest The authors declare no conflict of interest regarding this publication.

References

1. Zhang L, Lin J, Karim R (2016) Sliding window-based fault detection from high-dimensional data streams. IEEE Trans Syst Man Cybern Syst 47(2):289–303
2. Dai X, Gao Z (2013) From model, signal to knowledge: a data-driven perspective of fault detection and diagnosis. IEEE Trans Ind Inf 9(4):2226–2238
3. Gao Z, Cecati C, Ding SX (2015) A survey of fault diagnosis and fault-tolerant techniques-Part II: fault diagnosis with knowledge-based and hybrid/active approaches. IEEE Trans Ind Electron 62(6):3768–3774
4. Wang D, Peter WT (2015) Prognostics of slurry pumps based on a moving-average wear degradation index and a general sequential Monte Carlo method. Mech Syst Signal Process 56:213–229

5. Yin S, Kaynak O (2015) Big data for modern industry: challenges and trends [point of view]. Proc IEEE 103(2):143–146
6. Worden K, Staszewski WJ, Hensman JJ (2011) Natural computing for mechanical systems research: a tutorial overview. Mech Syst Signal Process 25(1):4–111
7. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828
8. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117
9. Wang M, Li H-X, Chen X, Chen Y (2016) Deep learning-based model reduction for distributed parameter systems. IEEE Trans Syst Man Cybern Syst 46(12):1664–1674
10. Y. LeCun, Y. Bengio, G. Hinton, and others, "Deep learning. nature 521 (7553), 436-444," Google Sch. Google Sch. Cross Ref Cross Ref, 2015
11. M. Gan, C. Wang, and others, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.,* vol. 72, pp. 92–104, 2016
12. Tamilselvan P, Wang P (2013) Failure diagnosis using deep belief learning based health state classification. Reliab. Eng. & Syst. Saf. 115:124–135
13. Chaturvedi I, Ong Y-S, Arumugam RV (2015) Deep transfer learning for classification of time-delayed Gaussian networks. Signal Process 110:250–262
14. Zellinger W, Moser BA, Grubinger T, Lughofer E, Natschläger T, Saminger-Platz S (2019) Robust Unsupervised Domain Adaptation for Neural Networks via Moment Alignment. Inf Sci 483:174–191
15. Nneji GU, Cai J, Jianhua D, Monday HN, Chikwendu IA, Oluwasanmi A, James EC, Mgbejime GT. Enhancing Low Quality in Radiograph Datasets Using Wavelet Transform Convolutional Neural Network and Generative Adversarial Network for COVID-19 Identification. *2021 4th Int Conf Pattern Recognit Artif Intell 2021*. p. 146–151. [https://doi.org/10.1109/PRAI53619.2021.9551043]
16. Nneji GU, Cai J, Jianhua D, Monday HN, Ejiyi CJ, James EC, Mgbejime GT, Oluwasanmi A. A Super-Resolution Generative Adversarial Network with Siamese CNN Based on Low Quality for Breast Cancer Identification. *2021 4th Int Conf Pattern Recognit Artif Intell 2021*. p. 218–223. [https://doi.org/10.1109/PRAI53619.2021.9551033]
17. Monday HN, Li JP, Nneji GU, James EC, Chikwendu IA, Ejiyi CJ, Oluwasanmi A, Mgbejime GT. The Capability of Multi Resolution Analysis: A Case Study of COVID-19 Diagnosis. *2021 4th Int Conf Pattern Recognit Artif Intell 2021*. p. 236–242. [https://doi.org/10.1109/PRAI53619.2021.9550802]
18. Gao Z, Cecati C, Ding SX (2015) A survey of fault diagnosis and fault-tolerant techniques-Part I: Fault diagnosis with model-based and signal-based approaches. IEEE Trans Ind Electron 62(6):3757–3767
19. E. J. Henley, "Application of expert systems to fault diagnosis," 1984
20. Shatnawi Y, Al-Khassaweneh M (2013) Fault diagnosis in internal combustion engines using extension neural network. IEEE Trans Ind Electron 61(3):1434–1443
21. Yin Z, Hou J (2016) Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. Neurocomputing 174:643–650
22. Lo CH, Fung EHK, Wong YK (2009) Intelligent automatic fault detection for actuator failures in aircraft. IEEE Trans. Ind. informatics 5(1):50–55
23. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX (2019) Deep learning and its applications to machine health monitoring. Mech Syst Signal Process 115:213–237
24. Cho HC, Knowles J, Fadali MS, Lee KS (2009) Fault detection and isolation of induction motors using recurrent neural networks and dynamic Bayesian modeling. IEEE Trans Control Syst Technol 18(2):430–437
25. Sun W, Shao S, Zhao R, Yan R, Zhang X, Chen X (2016) A sparse auto-encoder-based deep neural network approach for induction motor faults classification. Measurement 89:171–178
26. Lei Y, Jia F, Lin J, Xing S, Ding SX (2016) An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Trans Ind Electron 63(5):3137–3147
27. C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing,* vol. 130, pp. 377–388, 2017
28. Shao H, Jiang H, Zhang X, Niu M (2015) Rolling bearing fault diagnosis using an optimization deep belief network. Meas Sci Technol 26(11):115002
29. Jia F, Lei Y, Lin J, Zhou X, Lu N (2016) Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech Syst Signal Process 72:303–315
30. Lu J, Behbood V, Hao P, Zuo H, Xue S, Zhang G (2015) Transfer learning using computational intelligence: A survey. Knowledge-Based Syst. 80:14–23
31. Pan SJ, Yang Q (2009) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
32. Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. J. Big data 3(1):1–40
33. Pan SJ, Tsang IW, Kwok JT, Yang Q (2010) Domain adaptation via transfer component analysis. IEEE Trans Neural Networks 22(2):199–210
34. Liu C, Wu X, Jia Y (2015) Transfer latent SVM for joint recognition and localization of actions in videos. IEEE Trans. Cybern. 46(11):2596–2608
35. X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," 2011
36. C. Kandaswamy, L. M. Silva, L. A. Alexandre, R. Sousa, J. M. Santos, and J. M. de Sá, "Improving transfer learning accuracy by reusing stacked denoising autoencoders," in 2014 *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2014, pp. 1380–1387
37. Lu W, Liang B, Cheng Y, Meng D, Yang J, Zhang T (2016) Deep model based domain adaptation for fault diagnosis. IEEE Trans Ind Electron 64(3):2296–2305
38. K. A. Loparo, "Case western reserve university bearing data center," Bear. Vib. Data Sets, Case West. Reserv. Univ. http//csegroups. case. edu/bearingdatacenter/home, pp. 22–28, 2012
39. C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proceedings of the *IEEE conference on computer vision and pattern recognition,* 2017, pp. 4681–4690
40. X. Wang et al., "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops,* 2018, p. 0
41. N. C. Rakotonirina and A. Rasoanaivo, "ESRGAN+: Further improving enhanced super-resolution generative adversarial network," in ICASSP 2020-2020 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 2020, pp. 3637–3641
42. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision,* 2016, pp. 694–711
43. Sabour S, Frosst N, Hinton GE (2017) Dynamic routing between capsules. arXiv:1710.09829
44. G. E. Hinton, S. Sabour, and N. Frosst, "Matrix capsules with EM routing," 2018

45. Paoletti ME et al (2018) Capsule networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 57(4):2145–2160
46. Du W, Tao J, Li Y, Liu C. Wavelet leaders multifractal features based fault diagnosis of rotating mechanism. Mech Syst Signal Process Elsevier; 2014;43(1–2):57–75
47. Li Y, Wang X, Wu J. Fault diagnosis of rolling bearing based on permutation entropy and Extreme Learning Machine. 2016 Chinese Control Decis Conf 2016. p. 2966–2971
48. Zhang X, Liang Y, Zhou J, others. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. Measurement Elsevier; 2015;69:164–179
49. Haidong S, Hongkai J, Xingqiu L, Shuaipeng W (2018) Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine. Knowledge-Based Syst Elsevier 140:1–14
50. Gan M, Wang C, others. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. Mech Syst Signal Process Elsevier; 2016;72:92–104
51. Sohaib M, Kim C-H, Kim J-M. A hybrid feature model and deep-learning-based bearing fault diagnosis. Sensors Multidisciplinary Digital Publishing Institute; 2017;17(12):2876
52. Xu G, Liu M, Jiang Z, Söffker D, Shen W. Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. Sensors Multidisciplinary Digital Publishing Institute; 2019;19(5):1088