Turkish Journal of Electrical Engineering and Computer Sciences

Volume 30 | Number 5 Article 12

1-1-2022

A new automatic bearing fault size diagnosis using time-frequency images of CWT and deep transfer learning methods

YILMAZ KAYA

FATMA KUNCAN

HÜSEYİN METİN ERTUNÇ

Follow this and additional works at: https://journals.tubitak.gov.tr/elektrik

 Part of the Computer Engineering Commons, Computer Sciences Commons, and the Electrical and Computer Engineering Commons

Recommended Citation

KAYA, YILMAZ; KUNCAN, FATMA; and ERTUNÇ, HÜSEYİN METİN (2022) "A new automatic bearing fault size diagnosis using time-frequency images of CWT and deep transfer learning methods," *Turkish Journal of Electrical Engineering and Computer Sciences*: Vol. 30: No. 5, Article 12. https://doi.org/10.55730/1300-0632.3909

Available at: https://journals.tubitak.gov.tr/elektrik/vol30/iss5/12

This Article is brought to you for free and open access by TÜBİTAK Academic Journals. It has been accepted for inclusion in Turkish Journal of Electrical Engineering and Computer Sciences by an authorized editor of TÜBİTAK Academic Journals. For more information, please contact academic.publications@tubitak.gov.tr.

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2022) 30: 1851 – 1867
© TÜBİTAK
doi:10.55730/1300-0632.3909

A new automatic bearing fault size diagnosis using time–frequency images of CWT and deep transfer learning methods

Yılmaz KAYA[1], **Fatma KUNCAN**[1,*], **H. Metin ERTUNÇ**[2]
[1]Computer Engineering, Siirt University, Siirt, Turkey
[2]Mechatronics Engineering, Kocaeli University, Kocaeli, Turkey

Received: 15.10.2021 • **Accepted/Published Online:** 08.05.2022 • **Final Version:** 22.07.2022

Abstract: Bearings are generally used as bearings or turning elements. Bearings are subjected to high loads and rapid speeds. Furthermore, metal-to-metal contact within the bearing makes it sensitive. In today's machines, bearing failures disrupt the operation of the system or completely stop the system. Bearing failures that can occur can cause enormous damage to the entire system. Therefore, it is necessary to anticipate bearing failures and to carry out a regular diagnostic examination. Various systems have been developed for fault diagnosis. In recent years, deep transfer learning (DTL) methods are often preferred in current bearing diagnosis models, as they provide time savings and high success rates. Deep transfer learning models also improve diagnosis accuracy under certain conditions by greatly reducing human intervention. Diagnosis at the right time is very important for the sustainability and efficiency of industrial production. A technique based on continuous wavelet transform (CWT) and two dimensional (2D) convolutional neural networks (CNN) is presented in this paper to detect fault size from vibration data of various bearing failure types. Time-frequency (TF) color scalogram images for bearing vibration signals were obtained using the CWT method. Using AlexNet, GoogleNet, Resnet, VGG16, and VGG19 deep transfer learning methods with scalogram images, fault size prediction from vibration signals was performed. Five different transfer deep learning models were used for three different data sets. It was observed that the success rates obtained varied between 96.67% and 100%.

Key words: Bearing faults, fault detection, fault classification, continuous wavelet transform, convolutional neural networks, transfer learning

1. Introduction

Recent technological advancements have expedited the automation process by minimizing human participation in the sector and increasing the demand for new gear and equipment. Automation systems perform human-powered operations in a shorter time and are more stable. The growing usage of automation systems has expanded both the spare parts business and the maintenance and repair demands for these parts. Thus, maintenance-repair service has become essential for automation systems in the industry. The bearings, which are widely used in automation systems in the industry, are one of this equipment. Bearings are widely used in many machines and industrial applications. Bearings consist of an inner ring, an outer ring, a rolling element, and a cage. Bearings must be able to withstand immense mechanical stress. There are some cases where the bearing is damaged and the whole system is adversely affected. A malfunction that may occur in the bearing affects the whole system. In such cases, great financial losses are experienced in automation systems. For this reason, it is necessary to diagnose the bearings correctly and in advance. The no-fault operation of the bearing components is vital for the fault-free operation and life of the machines. A study from Europe shows that 34%

*Correspondence: fatmakuncan@siirt.edu.tr

This work is licensed under a Creative Commons Attribution 4.0 International License.

of the bearings could complete their life, and 66% of them were replaced early for various reasons. If changed early; it was seen that 16% could not complete their life due to assembly and disassembly errors, 36% could not complete their life due to lubrication errors, and 14% could not complete their life due to bad working conditions and pollution.

Faults in bearings are caused by mechanical friction, overheating, excessive or reverse loading, and contamination during operation. At the same time, the misalignment of the balls during production and the errors caused by the selection of nondurable materials causes the bearings to malfunction. These types of errors greatly affect the performance of the machines of which the bearings are a part. Early detection of bearing failures and taking measures against them have a crucial place in the efficient operation of the machines. A failure in bearings, which is one of the building blocks of mechanical transmission systems, greatly affects the stability, precision, and efficiency of the mechanism. If the relatively minor fault is ignored initially and allowed to grow further, it begins to threaten the safety and life of the mechanism. Therefore, these bearings, which have a very low cost compared to the mechanism, should be monitored correctly and the bearing must be replaced by stopping the system in case of failure [1–6].

Deep learning models are frequently preferred today in diagnostic models. On the other hand, deep learning techniques need a great quantity of labeled data. Because of this disadvantage, researchers prefer pretrained deep learning models. Therefore, the available features in the pretrained model are transferred to the transfer learning (TL) models to be used in a new problem. However, given the evolving new operating conditions in bearing failure analysis, collecting large amounts of labeled data for supervised diagnosis is costly and time-consuming. Therefore, when previously learned features are used in transfer learning models for a new problem, the data is not used in the training step and increases the effectiveness of the model in error diagnosis.

In this study, fault size estimation from bearing vibration signals was performed by using continuous wavelet transform (CWT) based scalogram images with transfer learning methods. 1D vibration signals have been converted into scalogram images using CWT methods. In the classification phase, transfer deep learning methods such as AlexNet, GoogleNET, Resnet, VGG16, and VGG19 were employed. Scalogram images are submitted as input to transfer methods. Using AlexNet, GoogleNet, Resnet, VGG16, and VGG19 deep transfer learning methods with scalogram images, fault size prediction from vibration signals was performed. Five different transfer deep learning was used for three different data sets. It was observed that the success rates obtained varied between 96.67% and 100%. Considering the obtained results, one can conclude that successful results were obtained with the proposed approach. The most essential contributions of this manuscript; the first are 3 original data sets of bearing vibration signals taken in different scenarios are used. The second one is that vibration signals are converted to scalogram images using CWT and these images are used as inputs to transfer deep learning methods. According to the results, it was observed that the proposed method provides a stable approach for detection bearing failures from vibration signals.

2. Literature studies

For many years, bearing fault analyses have been found worthy to research. Information about the bearing failure status can be obtained by collecting vibration signals in certain periods and analyzing these signals with an intelligent model. Many studies based on vibration signals of bearings were carried out using computer-based machine learning methods.

Arslan et al. [7] formed a theoretical model for the system by analyzing the regional surface defects in radial ball bearings using the vibration method and developed a computer program simulating the system due

to this model. McInerny and Dai [8] have performed signal envelope analysis, and the link between bearing failure frequencies and amplitude modulation/demodulation was explained. Orhan et al. [9] investigated the vibration behavior of a faulty bearing in a real system. Wang et al. developed a novel hybrid technique of the random forest classifier for rolling bearing diagnostics. The defect feature parameters were retrieved using wavelet packet decomposition, and the optimum primary wavelet set for signal preprocessing was defined using signal-to-noise ratio and mean square error values. Researchers stated that distinguish normal, inner, outer, and ball faulty bearings with an accuracy of 88.23% with the proposed method in the comparative experiment results [10]. Yan et al. proposed a new fault classification algorithm based on optimized SVM with multidomain features, including three phases (i.e. multidomain feature extraction, feature selection, and multidomain feature extraction). The researchers used three techniques (statistical analysis, fast Fourier transform (FFT), and variational mode decomposition (VMD)) independently in the first step to extract fault feature information from the multidomain direction (e.g., time domain, frequency domain, and time-frequency domain). They carried out feature selection with the Laplace score algorithm. Using the support vector machine classification model based on particle swarm optimization (PSO-SVM) to employ the detection of the bearings' various failure states, according to the results of the experiments, multidomain features give greater diagnostic accuracy than single-domain features [11]. Kuncan obtained bearing vibration signals from a special test setup and the signals were transferred to the 1D-LBP domain by applying this method to the vibration signals. Statistical features were obtained from the signals in the 1D-LBP domain and the vibration signals were classified by gray relational analysis (GRA) using these features [12].

Deep learning models, which are used in various fields in the literature, have started to show their effectiveness in bearing fault diagnosis studies. In this context, many researchers have started to study deep learning models for bearing land diagnoses. The fact that traditional methods take more processing time and the lack of a standard model has caused researchers to use transfer learning models more widely. Wang & Gao [13] first transformed the bearing vibration data into spectrum images with Morlet-based wavelet transform. Later, researchers classified the bearings as normal, inner, outer ring, ball defective bearing with the VGG19 transfer learning model. They achieved 93.9% success with this method. Mao et al. realized the classification of the normal, initial stage, and heavy faulty bearings by using the VGG16 transfer learning model. The researchers created a three-channel data set by combining TF domain information. Original vibration data, Hilbert–Huang transform (HHT), and fast Fourier transform (FFT) transformations of vibration data were merged to form a 3D dataset. They stated that the model they created determined the starting point of the bearing failure without any errors. In addition, comparative studies have shown the superiority of the transfer learning model over traditional methods [14]. Shao et al. used the VGG16 model in three datasets containing faults in induction motors, gearboxes, and bearings. First, they converted the data set they obtained from these three experimental sets into TF images. Later, they classified these images using the VGG16 model. In the data set of induction motor, the authors achieved to classify the data set consisting of normal motor, stator winding faulty motor, unbalanced rotor motor, bearing defective motor, broken bar motor, inclined rotor motor with 100% accuracy. They achieved 99.31% success in the classification of the gearbox dataset (worn, missing parts, welding faulty, surface faulty types). For the last dataset of bearing, they succeed 99.82% accuracy in the classification of normal, outer ring, inner ring, ball defective ball bearing, and both inner and outer ring faulty bearings [15]. Wang et al. converted vibration data into spectrogram images using FFT. In their studies, researchers obtained low-level properties of ResNet-50 by using the Resnet-50 model and created a new feature set by replacing the last two layers of the model with the multiscale feature extractor. They tested their models

using the public data set (Case Western Reserve University–CWRU) bearing data set and the data set which created by them in their test setup. In the CWRU dataset, the researchers achieved the highest success in the classification of IRF, ORF, and BF defective bearings under different operating conditions, 99.80%, and in the datasets, 99.23% in the classification of IRF, ORF, and BF defective bearings. Ma et al. transformed the vibration data obtained from the CWRU data set into 2D images with three different methods. These images are obtained by frequency slice, wavelet transforms, continuous wavelet transform, and reconstructed image methods, respectively. They distinguished the types of bearing faults using the Alexnet architecture and applied t-distributed stochastic neighbor embedding (t-SNE) to visualize the feature learning process. The researchers achieved a 99.19% success rate with this model which was developed by them [16]. Lu et al. used Alexnet to identify the different types of bearing failure in the CWRU dataset. To test the effectiveness of the suggested approach, they choose five bearing faults obtained under two different motor loads and speeds. Then, the researchers transformed the vibration signals into spectrogram images with the nonuniform FFT method and they could distinguish the bearing faults types with a 99.7% success rate [17]. Wen et al. used the ResNet-50 model with 51 convolutional layer depth for fault diagnosis in their studies. Firstly, they converted the vibration signals into gray images. They tested the model they presented in three different data sets. They stated that they achieved 98.95% success for the Paderborn KAT bearing data set, 99.99% for the CWRU data set, and 99.20% for the self-priming centrifugal pump data set [18]. As a conclusion of the studies discussed, the researchers have usually obtained prominent results for diagnosing bearing faults based on deep learning techniques.

3. Datasets

In this study, 3 data sets taken in different scenarios are used. Information about the data sets is given below.
Data Set 1 (DS1): It consists of the signals obtained at the same speed (2100 rpm) test setup of the bearings with fault-free (0 mm), 0.25 mm outer ring fault, 0.5 mm outer ring fault, and 0.75 mm outer ring fault. The data obtained in this data set were obtained by running the fault types of different fault sizes at 2100 rpm in a certain time period in the experimental setup.
Data Set 2 (DS2): This dataset consists of the signals obtained at the same speed (2100 rpm) test setup of bearings with a fault-free (0 mm), 0.25 mm inner ring fault, 0.5 mm inner ring fault, and 0.75 mm inner ring fault. The data obtained in this data set were obtained by running the error types of different error sizes at 2100 rpm in a certain time period in the experimental setup.
Data Set 3 (DS3): The dataset consists of 3 faults (0 mm), 0.25 mm ball fault, 0.5 mm ball fault, and 0.75 mm ball fault, obtained from the same speed (2100 rpm) test setup. The data obtained in this data set were obtained by running the error types of different error sizes at 2100 rpm in a certain time period in the experimental setup.

Figure 1 shows the setup of the study. In this setup, the servo motor from which the data will be taken, the accelerometer sensors that serve as an acceleration measurement, as well as the bearing holding mechanism in the servo motor, the tool kit for various mechanical disassembly operations, the computer that provides the analysis and processing of the data, the valve and the radial valve in the system that provides the movement of the piston load. In addition, there is the I/O driver that performs the opening and closing operations of the system and the DAQ Card unit that collects the data.

The working principle of this system is to follow the resulting scenarios by sending artificial errors to the servo motor. Based on these scenarios, a precautionary mechanism can be created against errors.

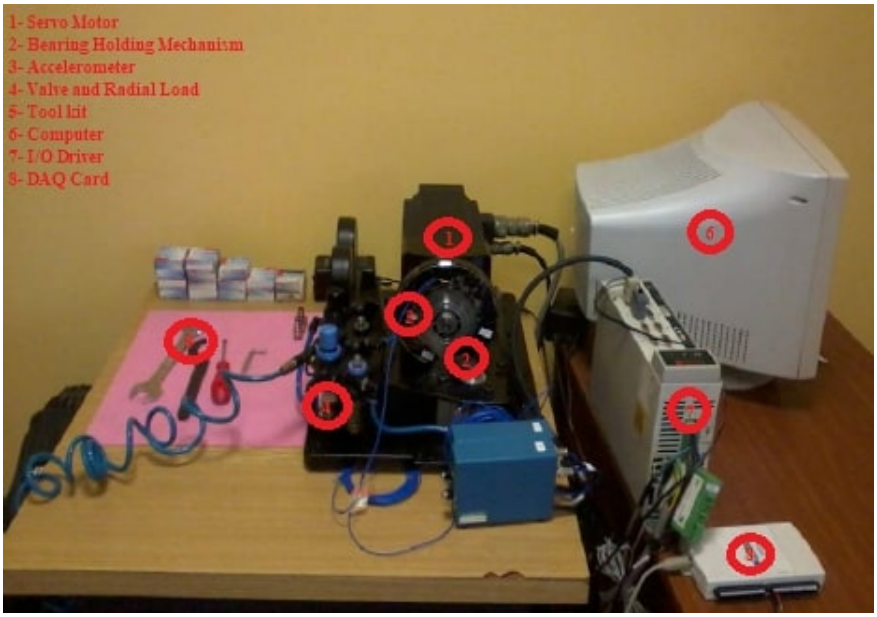


Figure 1. Setup and components.

4. Methods

4.1. Continuous wavelet transform

In the frequency and time domains, the wavelet transform is suitable for representing temporal characteristics locally. A wavelet transform is a function that combines a wavelet with a time series of a function. Continuous wavelet transform (CWT) is used to generate a frequency and time function scalogram to achieve better time localization for short duration high-frequency events and better frequency localization for low-frequency long-duration events. A scalogram is the absolute value of the coefficients obtained using CWTs of a signal. It depicts how much each frequency band component contributes to the energy of the signal over time intervals [19–20].

CWT helps to distinguish noise from signals by having TF representation. It is difficult to distinguish noise from signals simply by the time domain. Wavelets are useful in analyzing such signals. The wavelet transform is excellent at representing temporal properties locally in both time and frequency domains. Wavelet transform is a joint function of wavelet and time series of a function. CWT is used to produce fine and common TF analysis and to help localize frequency information.

The TF diagram is obtained from the analysis of signals whose frequency varies according to time with CWT. The method chosen for TF domain transformation is crucial in pattern recognition applications. Wavelet transform is a suitable method for such applications. Literature has shown that CWT is an effective method for nonstationary signals such as EEG, EMG, and ECG. In wavelet transformation, wavelet functions such as Daubechies, Morlet, Symlets, and Gaussian are employed. In this study, Morlet wavelet transform, which is widely used, was preferred. It is known that the wavelet is a small wave-like function of limited duration. The family of $\psi_{(s,\tau)}$ (t) functions is defined as the version of a single main wavelet function $\psi(t)\epsilon L^2$ (R) transformation parameter ($\tau\epsilon R$) and expansion parameter (sϵR) [21].

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \int \psi(\frac{t-\tau}{s}) \tag{1}$$

Here $\psi_{(s,\tau)}$ (t) denotes the one-dimensional signal being converted. τ is the wavelet function; s is the transform parameter and ψ is the main wavelet function. When the scale parameter (s) is large, it expands the signal and is used at low frequencies; the small-scale parameter (s) compresses the signal and is used at high frequencies. Here, $(1/\sqrt{|s|})$ normalization states that $\parallel \psi_{(s,\tau)}(t) \parallel$ is independent of s and τ parameters. Wavelet functions are usually normalized to have a unit $(\parallel \psi_{(s,\tau)}(t) \parallel = 1)$. The CWT is calculated as an inner product of $x(t)\epsilon L^2(R)$ with the translation and expansion of normalized wavelets as $\frac{1}{\sqrt{s}}\psi(\frac{t-\tau}{s})$. With the CWT, the coefficients of the signals are calculated with the following equation [22–24].

$$W_x^{\psi}(s,\tau) = \langle x(t), \psi_{s,\tau} \rangle \int_{-\infty}^{\infty} x(t)\psi^*(\frac{t-\tau}{s})dt \tag{2}$$

The wavelet transform coefficients are generated by associating the normalized wavelet function at various scales and sliding over the data at any point in time. CWT not only affects the expansion of the mother wavelets but also the amplitude. When the parameter s is high, the wavelet amplitude is low, and when s is low, the wavelet amplitude is high. The wavelet function scans the input signal x based on the transformation parameter for a value of a given scale parameter, and this process is repeated for each value of a parameter. The scalogram images utilized in the study are a measure of the energy distribution in the signal's time shift ( ) and scaling factor (s).

A scalogram is the absolute value of the coefficients obtained using CWTs of a signal. It shows how much each component of the frequency domain contributes to the energy of the signal over time intervals. The energy formula is defined by the following equation [22,24].

$$Energy_{cwt}(\tau, s) = \sum_{\tau}\sum_{s}(X_{cwt}(\tau, s))^2 \tag{3}$$

4.2. Convolutional neural network (CNN)

Convolutional neural network (CNN) was motivated by the biological regulation of the visual cortex with simple and complex cells [25–26]. Convolutional neural network has three major characteristics that make them appropriate for extracting features from pictures. Local receptive fields, weight sharing, and spatial subsampling are all essential aspects [27]. Convolutional neural network is a multilayer feedforward NN that includes different layer types with convolution layers, ReLU layer, pooling layers, feature extraction layers with batch normalization, and fully connected output layers. Convolutional neural network (CNN) is designed to recognize micro-macro features, including edges and shapes in images. The layers in convolutional neural network (CNN) are briefly summarized below. The first layer is the input layer to a convolutional neural network (CNN) that accepts M × N × 3. The convolution layer is the initial layer that extracts information from an input image by convolutional an image with a filter or kernel.

Convolutional neural network (CNN) employs a filter with particular parameters with the same depth as the input picture, which is then merged with the image. The filter represents a curve or shape that joins the input image. The shape resembling the curve in the input image represented by the filter reaches higher values because of convolution. A 3D feature map in the size of $(\frac{a-c}{stride}+1)x(\frac{b-d}{stride}+1)x(n)$ is obtained when n filters (f) of size (c × d × 1) are applied to an a × b × 1 gray image (I). When the number of steps (stride) is

1 (one), then the output will be (a–c+1) × (b–d+1) × (n). The convolution process between filter vectors can be described mathematically as follows [28]:

$$X_j^l = \vartheta(\sum_c X_i^{l-1} * f_{cd}^l + b_j^l). \tag{4}$$

Here ϑ is the activation function, f is the filter matrix and l is the l^{th} layer in the network [28].

Pooling layer: When images are too large, pooling layers can be used to reduce the number of parameters. Pooling is the process of grouping matrix data into various segments and replacing the entire segment with a single value, resulting in a reduction in the amount of metric data. Some of the popular pooling functions are max pooling and average pooling. When e × f × 1 filter is applied to the image whose output is (a–c + 1) × (b–d + 1) × (n) of the convolution layer, the image is ((a–c + 1) – e + 1) × ((b–d + 1) – f + 1) transforms into an image of size x(n). The depth of the image is preserved. However, there is a decrease in the two dimensions of the image. Mathematically, the pooling process can be defined as follows:

$$x_j^l = \vartheta(\beta_j^l down(z_i^{l-1} + b_j^l). \tag{5}$$

Rectified linear unit (RELU) Layer: Values less than zero in this layer are set to zero. It can be mathematically expressed as follows [21]:

$$f(z) = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases} \tag{6}$$

Batch normalization layer: The main function of this layer is to normalize along each input channel. This layer is provided to increase training speed and reduce sensitivity to network startup.

Fully connected layer (FCL): The output of the pooling layer is converted to a one-dimensional vector and then given to the FCL. An FCL connects all output from the previous layer to the next layer of the network, as is usually the case with a traditional neural network. The input to this layer is multiplied by a weight matrix and summed by the threshold value.

Softmax layer: This layer is the activation function layer. The softmax function transforms the preceding layers' input into a probability for classes with a total of one. As a result, this layer is essential in the output since the projected output is the class with the max. probability for the given input data [21].

Classification output layer: Cross-entropy loss values are calculated in this layer. The layer estimates the number of classes based on the preceding layer's output size. This layer takes the softmax function input and collectively assigns each input to one of its classes using the cross-entropy loss function, which is theoretically described by the following equation.

$$loss = \sum_{v=1}^{Nu} \sum_{w=1}^{C} s_{vw} ln y_{vw} \tag{7}$$

Here Nu is the sample size and C is the number of classes. S_{vw} is an indicator to define the v_{th} instance belongs to the wth class. Y_{vw} is output for example v for class w.

By combining the aforementioned layers in various ways, many distinct deep learning methods are formed [22, 29, 30].

4.3. Deep transfer learning methods

The convolution neural network is rarely trained from the ground up. Instead, it is standard practice to utilize a pretrained model trained on big datasets such as ImageNet and then apply the pretrained weights to the situation at hand. The pretrained network may be used as a fixed feature extractor from the input image, or it can be used to fine-tune the entire network or a part of it. The pretrained model is employed as a fixed feature extractor when the dataset is small and similar to the ImageNet dataset. When the dataset is big and similar to the ImageNet dataset, not only is the final layer of the network eliminated but the weights of the whole network (or portion of the network) are changed using the backpropagation technique to improve accuracy. When the dataset is big and does not match the ImageNet dataset, the whole network is trained from scratch or using the ImageNet classifier's pretrained weights. AlexNet, ResNet, GoogleNEt, Vgg16, and Vgg19 are some pretrained convolutional neural network (CNN) architectures for transfer learning [21].

Transfer deep network models will be used in this study. Although these networks have been pretrained to classify pictures, we can use transfer learning to adapt them to match our classification problem by changing the needed parameters. The weights taken from pretrained deep convolutional neural network (CNN) models are used in TL. The structure of the TL technique used to detect the scalogram images obtained for the vibration signals of different fault sizes is given in Figure 2. Firstly, 1D vibration signals are transformed into scalogram images. These images are submitted as input to these pretrained deep learning models. The weights and hyperparameters of the pretrained model are transferred from the pretrained model, and the network or group of networks added to the end of the model is trained using the specified hyperparameters. Training of the entire model is obtained after these procedures [22].

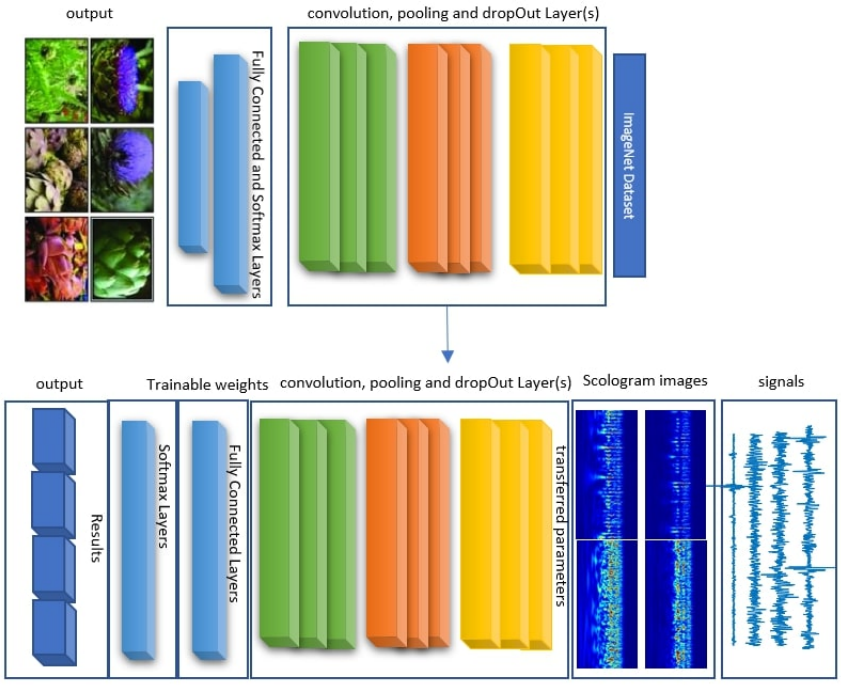


Figure 2. Structure of the transfer deep learning method for fault size prediction.

4.3.1. Vgg16 and Vgg19

Simonyan and Zisserman proposed similar Vgg16 and Vgg19 methods in 2014 [31]. The most crucial feature of these models was to utilize only small convolution filters (3×3). With the capability of the GPU cluster, it was possible to expand the depth of the network to 16 and 19 layers. Because of its simple design and strong generalization capabilities, the VGG model is now regarded as one of the finest TL approaches used in image recognition applications. The VGG-16 model comprises around 134 million parameters and 16 trainable layers, which include convolutional and fully connected, maximum pooling, and dropout layers. VGG-19 contains 144 million parameters and 19 trainable layers.

4.3.2. ResNet

Transfer methods are developed on the ImageNet dataset. And new models are emerging with more layer structures. Thus, the depth of models increases. However, as the model deepens, it gets intricate to train. Common problems that arose are breakdown and gradients that disappear (at some point the accuracy saturates and then drops rapidly). A residual learning framework in deep learning networks aids in the maintenance of excellent outcomes via a neural network with several layers. ResNet is made up of layers that are stacked on top of each other. This design is differentiated by the usage of a shortcut in each layer to connect the input straight to the output [32]. The layer is now referred to as a block as a result of this shortcut. This network has shown to be the most effective model. for the classification problem organized for the ImageNet dataset in 2015, offering 152 layered models. Despite being eight times deeper than the VGG-19, it still has lower complexity.

4.3.3. GoogleNet

Szegedy et al. granted the inception architecture. This example is called as GoogLeNet [33]. It performed significantly better than the VGG-19 while having 12 times fewer parameters. The objective (and the most obvious way to improve the performance of the network) was to increase its size in terms of both depth and width. However, this simple method has serious drawbacks, such as the possibility of memorization and, more critically, the enormous increase in processing resources required. The proposed method was to transition from fully connected to sparsely connected systems. As a result, the Inception model's primary principle is to combine multiple layers into one sort of block in parallel rather than stacking them on top of each other. It was assumed that a network using such an approach would select the most useful layers that increase their weight and at the same time reduce useless layers. In addition, 1×1 convolution filters are used to help reduce the size of the feature maps and global average pooling.

4.3.4. AlexNet

This architecture, designed by Alex Krizhevsky, is one of the first networks to make ImageNet classification accuracy a significant step forward compared to traditional methods. This model was introduced in 2012 and showed better performance than the previous latest technology solutions. It benefited from the GPU to make the convolution process more rapid and more efficient [30]. Because computational resources are the major constraint, the architecture has been optimized to utilize two current GPUs for parallel computations. AlexNet is a basic model that consists of five convolutional layers followed by maximum pooling layers, which are utilized together for feature extraction. The network employs three fully linked layers with Softmax activation for the classification process. For improved training performance, nonsaturating ReLU activation functions are utilized.

5. Experimental results

Wavelet transform methods like CWT only work at a single scale, that is, neither time nor frequency. It works as a multiscale transformation that splits the signal into multiple scales. A scalogram utilized in this study, which is represented by the absolute values of the signal's CWT and may be displayed as a function of both frequency and time. One-dimensional vibration signals are converted into images through CWT using a set of filters. These images contain the TF and amplitude information of the signal. The filter set specifies the CWT parameters to be applied to the given signal. In this study, filter parameters are kept constant for CWT. Since the wavelet used for CWT has better TF localization, the analytical Morse wavelet has been used [34]. Deep neural networks use convolutional neural network (CNN) for feature recognition. With CWT, vibration signals are transformed into equivalent images and made suitable for training deep networks. The image after CWT includes time, scale, and amplitude information. The produced 2-D picture is rescaled in the last preprocessing phase according to the input requirement of the specific deep web, for example, Alexnet has an input size of 224 x 224 pixels. As a result, the image was scaled to 224 x 224 pixels before being put into Alexnet for transfer learning. While smaller data sets are sufficient for training traditional artificial neural networks, deep networks require a large sample size for training. For this purpose, transfer deep networks are preferred for modeling small data sets. Because transfer deep networks are pretrained networks [35–39]. When these networks are used in small data sets trained for different large data sets, only some layers of these networks are retrained. Transfer deep learning methods are designed to classify images. To classify the vibration signals, the signals must be converted into images beforehand. In this study, vibration signals were transformed into images called scalograms with CWT. GoogleNet, AlexNet, ResNet50, Vgg16, and Vgg19 deep transfer methods were used for the classification of the images. Each of the deep networks used has a size requirement for the input image. The images are resized according to the model used. In this investigation, three separate data sets were used. Each data set includes vibration signals for various faults. These signals have been converted into images. A total of 3800 images were obtained for each data set. There are four classes for each data set. Seventy percent of each data set was used for training (2660 images) and 30% (1140 images) for testing. The model parameters for all networks are kept constant. The training process was carried out with the same parameter values. To avoid memorization problems, the number of epochs was allowed to reach a maximum of 25. The learning rate was used as 0.0001. In the trials, the Adam optimizer and the categorical cross-entropy loss function were used. Scalogram images were created by using sample signals in 3 different data sets. Images of these sample signals are given in Figures 3–5. As can be seen from the figures, it is seen that the scalogram images differ from each other for each data set. Transfer deep learning success rates of data sets are given in Table 1.

Five different transfer deep learning methods were used for training-test data sets of 3 data sets. A high success rate of 100% was achieved for DS2 and DS3. The lowest success rate for DS3 is 99.11% with GoogleNet. For DS2, the lowest success rates were obtained with GoogleNet and AlexNet as 98.21%. Lower success rates were observed for DS1 compared to other DS3 and DS2 data sets. For this data set, the highest success rates were obtained with VGG16 and VGG19 as 99.12%. The lowest success rate with GoogleNet was 96.67%.

Since 100% success rates are generally observed for DS2 and DS3, the success and loss values of GoogleNet for the DS1 data set are given in Figure 6. The model with the lowest performance for all data sets was observed as GoogleNet. Confusion matrices of GoogleNet for DS1, DS2, and DS3 are shown in Figure 7.

Some studies for bearing failure prediction are given in Table 2. The performances of the studies were compared with the results obtained in this study. Looking at the table, it is understood that an acceptable result is achieved with the proposed approach.

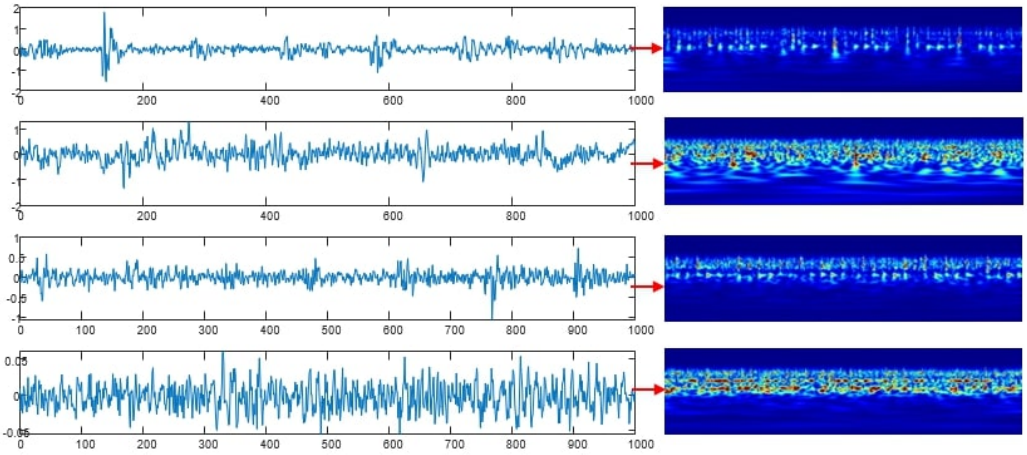


Figure 3. Scalogram displays of sample signals for DS1.

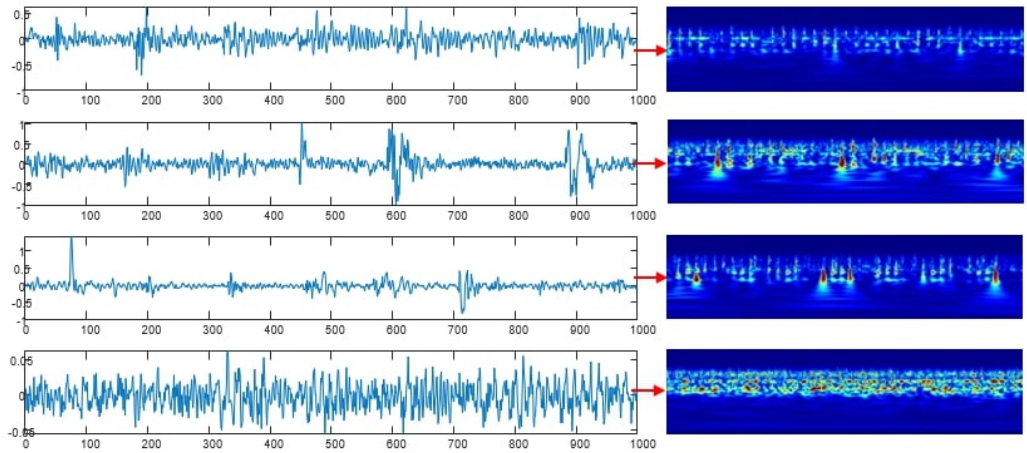


Figure 4. Scalogram displays of sample signals for DS2.

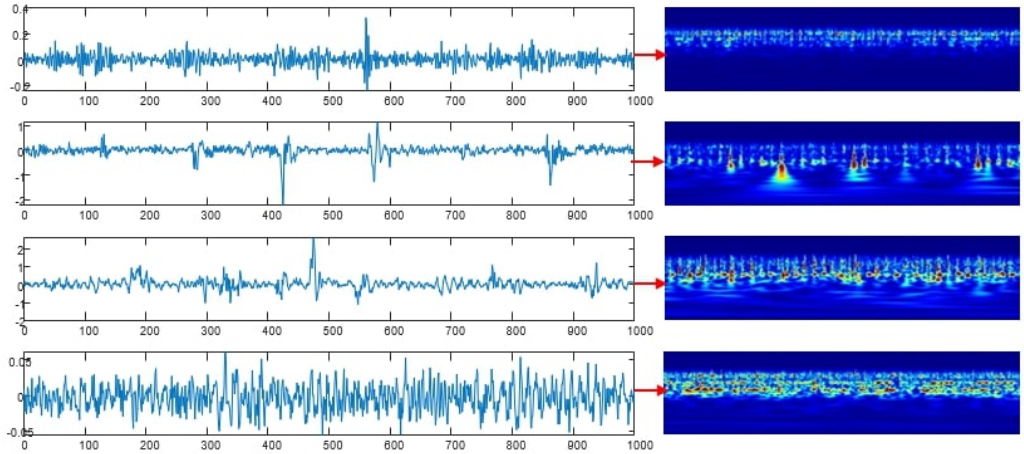


Figure 5. Scalogram displays of sample signals for DS3.

Table 1. Classification accuracy using the test data set.

Data set	**Model**	**Accuracy (%)**	Kappa score
Dataset 1	GoogleNet	96.67	0.9556
	AlexNet	98.21	0.9761
	ResNet50	98.25	0.9767
	Vgg16	99.12	0.9883
	Vgg19	99.12	0.9883
Dataset 2	GoogleNet	98.21	0.9761
	AlexNet	98.21	0.9761
	ResNet50	100	1
	Vgg16	100	1
	Vgg19	100	1
Dataset 3	GoogleNet	99.11	0.9881
	AlexNet	100	1
	ResNet50	100	1
	Vgg16	100	1
	Vgg19	100	1

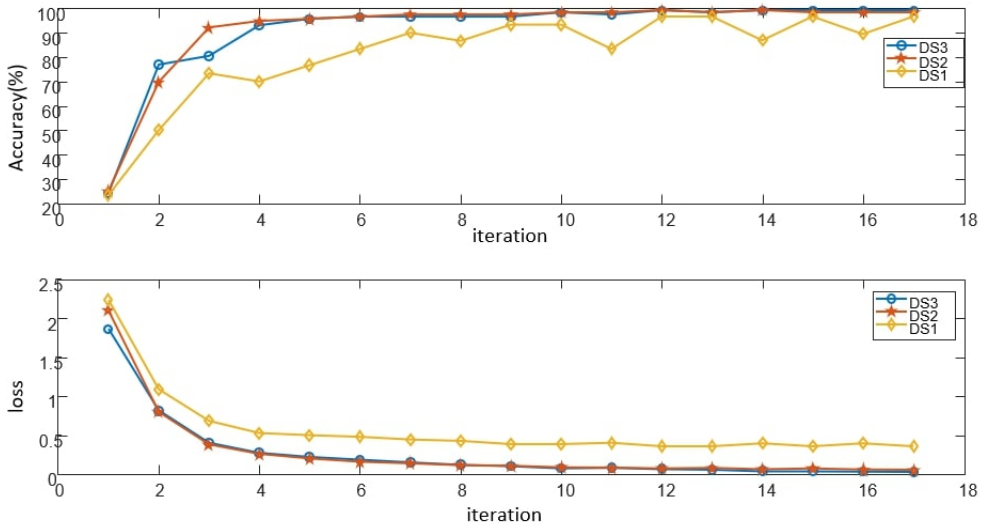


Figure 6. Accuracy and Loss Values for DS1with GoogleNet.

6. Discussion

In the machines used in today's automation systems, the movement is mostly carried out by rotational force. Bearings are rolling machine parts that are frequently utilized in motor systems that produce this rotating motion. Various faults may occur on the bearing parts during manufacturing or assembly, and the continuous mechanical friction of the bearings will make it inevitable to wear over time. Machine operating conditions and ambient structure may potentially result in bearing failures. When a faulty bearing is not replaced in time, it can cause irreversible damage to the workpiece or the system to which it is attached. The bearing replaced prematurely causes to stop production unnecessarily. It is undesirable to stop production because of

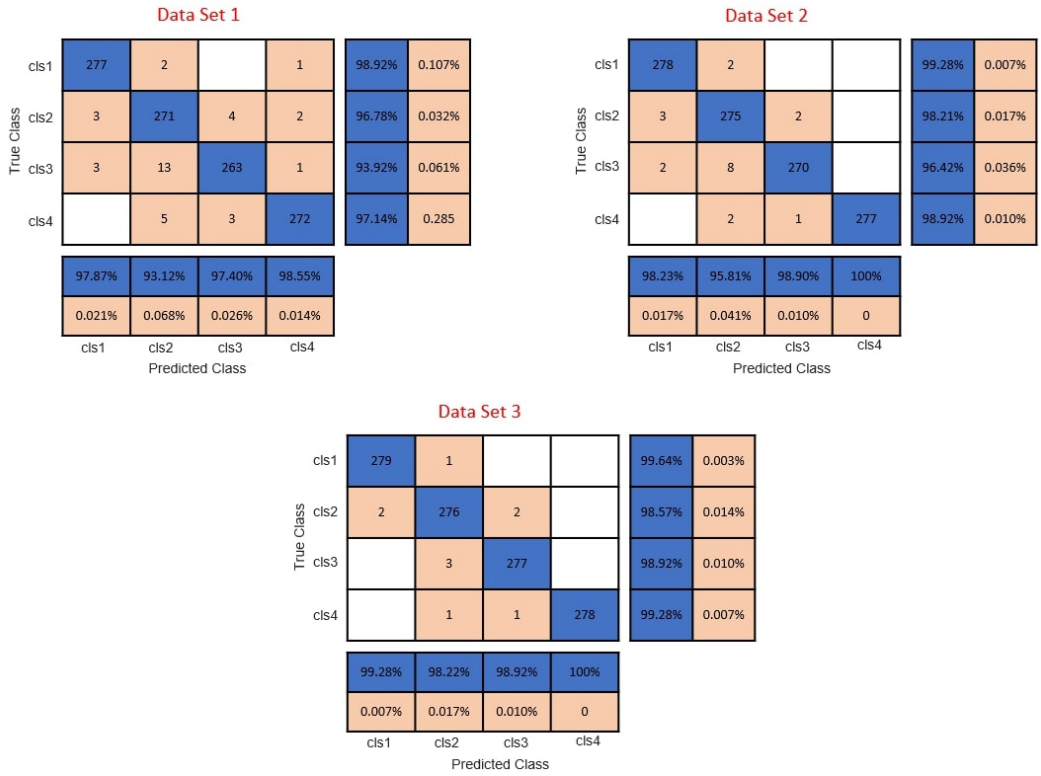


Figure 7. Confusion Matrices for accuracy 96.67% for DS1, 98.21% for DS2 and 99.11% for DS3 using GoogleNet.

both malfunction and faulty fault detection. For this reason, bearing faults must be detected before they reach a dangerous level and necessary precautions must be taken.

Information about the bearing failure status can be obtained by collecting vibration signals in certain periods and analyzing these signals with an intelligent model. Bearing fault analysis has been an area of interest for researchers for many years. In particular, analysis studies of vibration signals of bearings were carried out using computer-based machine learning methods. Deep learning models, which are used in various fields in the literature, have started to show their effectiveness in bearing fault diagnosis studies. Recently, many researchers have started to study deep learning models for bearing fault diagnoses. High success rates can be observed in transfer deep learning methods even if the data set is limited. The only disadvantage of such methods is that the training process is longer than traditional methods. However, after the models are trained, the classification process will be fast.

In this study, a method based on continuous wavelet transform (CWT) and two-dimensional (2D) convolutional neural networks (CNN) is proposed to determine fault size from vibration signals of different fault types in bearings. A deep learning-based method was used that uses the TF spectrum of bearing vibration signals automatically and uses single-channel vibration signals without the requirement for extraction of features. TF color scalogram images for bearing vibration signals were obtained using the CWT method. Using AlexNet, GoogleNet, Resnet, VGG16, and VGG19 deep transfer learning methods with scalogram images, fault size prediction from vibration signals was carried out. Five different transfer deep learning was used for three different data sets. It was observed that the success rates obtained varied between 96.67% and 100%. In our future studies, different feature extraction approaches or classification methods will be used in the classification

Table 2. Comparison of the proposed approach with studies in the literature.

Author/Year	Method	DataSet	Fault type	Accuracies (%)
Ke Li et al., 2017 [40]	- Kernel extreme learning machine (KELM) - PSO-KELM	- Authors's setup	- EFB, ORF, IRF, and BF	80%–95%
Ju et al., 2018 [41]	- Improved multi scale entropy (IMSE)	- Case Western Reserve University (CWRU) Bearing Data Center	- Error free bearing (EFB), inner ring fault (IRF), outer ring fault (ORF), and ball fault (BF)	95%–98.57%
Ge et al., 2019 [42]	- Kernel principal component analysis (KPCA)	- CWRU Bearing Data Center	- Error free bearing (EFB), inner ring fault (IRF), outer ring fault (ORF), and ball fault (BF)	78.75%–95%
Zhu et al., 2019 [43]	- Multiscale fuzzy measure entropy (MFME)	- CWRU Bearing Data Center	- Error free bearing (EFB), inner ring fault (IRF), outer ring fault (ORF), and ball fault (BF)	90.95%–100%
Kuncan, 2020 [12]	- 1D-LBP+GRA	- Author's setup	- Error free bearing (EFB), inner ring fault (IRF), outer ring fault (ORF), and ball fault (BF)	99%–100%
Shao et al., 2018 [15]	- Deep transfer learning	- CWRU Bearing Data Center	- Error free bearing (EFB), inner ring fault (IRF), outer ring fault (ORF), and ball fault (BF)	94.8% - 99.64%
Wang et al., 2020 [13]	-Deep learning (DL)	- CWRU Bearing Data Center	Fault type -Normal, inner-race, outer-race, and ball Fault severity (defect size, mm) - 0, 0.18, 0.36, and 0.54	93%
Wang et al., 2020 [16]	- Multiscale deep intraclass transfer learning	- CWRU Bearing Data Center	- Error free bearing (EFB), inner ring fault (IRF), outer ring fault (ORF), and ball fault (BF)	99%
Wen et al., 2020 [18]	- ResNet-50	- CWRU Bearing Data Center	- Error free bearing (EFB), inner ring fault (IRF), outer ring fault (ORF), and ball fault (BF)	98.95%–99.99% and 99.20%
This Study	CWT+Transfer deep learning methods	Authors's setup	- EFB, ORF, IRF, BF	96.67%–100%

of bearing vibration signs or in fault detection. Bearing vibration signals will be obtained at different speeds, failures in different regions, and different failure rates, and experiments will be carried out.

Acknowledgment

This work was supported by the Scientific Research Projects Coordination Unit of Siirt University as a project with the number 2021-SİÜMÜH-14. This study was performed in Siirt University Faculty of Engineering Machine Vision (MaVi) Laboratory. The authors would like to thank the staff of MaVi Laboratory for their support.

References

[1] Xu Y, Tian W, Zhang K, Ma C. Application of an enhanced fast kurtogram based on empirical wavelet transform for bearing fault diagnosis. Measurement Science and Technology 2019; 30 (3): 035001.

[2] Kuncan M, Kaplan K, Minaz MR, Kaya Y, Ertunc HM. A novel feature extraction method for bearing fault classification with one dimensional ternary patterns. ISA transactions 2020; 100: 346-357.

[3] Wu Z, Jiang H, Zhao K, Li X. An adaptive deep transfer learning method for bearing fault diagnosis. Measurement 2020; 151: 107227.

[4] Kaya Y, Kuncan M, Kaplan K, Minaz MR, Ertunç HM. A new feature extraction approach based on one dimensional gray level co-occurrence matrices for bearing fault classification. Journal of Experimental & Theoretical Artificial Intelligence 2021; 33 (1): 161-178.

[5] Ma P, Zhang H, Fan W, Wang C, Wen G et al. A novel bearing fault diagnosis method based on 2D image representation and transfer learning-convolutional neural network. Measurement Science and Technology 2019; 30 (5): 055402.

[6] Kaplan K, Kaya Y, Kuncan M, Mînaz MR, Ertunç HM. An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis. Applied Soft Computing 2020; 87: 106019.

[7] Arslan H, Orhan S, Aktürk N. "Bilyalı Rulman Hasarlarının Neden Oldugu Titresimlerin Modellenmesi", Gazi Üniv. Müh. Mim. Fak. Der. 2003; 18 (4): 123- 146.

[8] McInerny SA, Dai Y. "Basic Vibration Signal Processing for Bearing Fault Detection", IEEE Transactions on Education 2003; 46 (1).

[9] Orhan S, Arslan H, Aktürk N. "Titresim Analiziyle Rulman Arızalarının Belirlenmesi", Gazi Üniv Müh. Mim. Fak. Der. 2003; 18 (2):39-48.

[10] Wang Z, Zhang Q, Xiong J, Xiao M, Sun G et al. Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests. IEEE Sensors Journal 2017; 17 (17): 5581-5588.

[11] Yan X, Jia M. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. Neurocomputing 2018; 313: 47-64.

[12] Kuncan M. An intelligent approach for bearing fault diagnosis: combination of 1D-LBP and GRA. IEEE Access 2020; 8: 137517-137529.

[13] Wang P, Gao RX. Transfer learning for enhanced machine fault diagnosis in manufacturing. CIRP Annals 2020; 69 (1): 413-416.

[14] Mao W, Ding L, Tian S, Liang X. Online detection for bearing incipient fault based on deep transfer learning. Measurement 2020; 152: 107278.

[15] Shao S, McAleer S, Yan R, Baldi P. Highly accurate machine fault diagnosis using deep transfer learning. IEEE Transactions on Industrial Informatics 2018; 15 (4): 2446-2455.

[16] Wang X, Shen C, Xia M, Wang D, Zhu J et al. Multi-scale deep intra-class transfer learning for bearing fault diagnosis. Reliability Engineering & System Safety 2020; 202: 107050.

[17] Lu T, Yu F, Han B, Wang J. A Generic Intelligent Bearing Fault Diagnosis System Using Convolutional Neural Networks With Transfer Learning. IEEE Access 2020; 8: 164807-164814.

[18] Wen L, Li X, Gao L. A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Computing and Applications 2020; 1-14.

[19] Garg D, Verma GK. Emotion recognition in valence-arousal space from multi-channel EEG data and wavelet based deep learning framework. Procedia Computer Science 2020; 171: 857-867.

[20] Silva A, Zarzo A, González JMM, Munoz-Guijosa JM. Early fault detection of single-point rub in gas turbines with accelerometers on the casing based on continuous wavelet transform. Journal of Sound and Vibration 2020; 487: 115628.

[21] Jadhav P, Rajguru G, Datta D, Mukhopadhyay S. Automatic sleep stage classification using time–frequency images of CWT and transfer learning using convolution neural network. Biocybernetics and Biomedical Engineering 2020; 40 (1): 494-504.

[22] Narin A. Detection of Focal and Non-focal Epileptic Seizure Using Continuous Wavelet Transform-Based Scalogram Images and Pre-trained Deep Neural Networks. IRBM 2020.

[23] Garg D, Verma GK. Emotion recognition in valence-arousal space from multi-channel EEG data and wavelet based deep learning framework. Procedia Computer Science 2020; 171: 857-867.

[24] Kaya D. The mRMR-CNN based influential support decision system approach to classify EEG signals. Measurement 2020; 156: 107602.

[25] Chen Z, Gryllias K, Li W. Mechanical fault diagnosis using Convolutional Neural Networks and Extreme Learning Machine. Mechanical Systems and Signal Process 2019; 133.

[26] Zhang W, Li C, Peng G, Chen Y, Zhang Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Sig. Process 2018; 100: 439–453.

[27] Kumar A, Zhou Y, Gandhi CP, Kumar R, Xiang J. Bearing defect size assessment using wavelet transform based Deep Convolutional Neural Network (DCNN). Alexandria Engineering Journal 2020; 59 (2): 999-1012.

[28] Kant P, Laskar SH, Hazarika J, Mahamune R. CWT Based Transfer Learning for Motor Imagery Classification for Brain computer Interfaces. Journal of Neuroscience Methods 2020; 345: 108886.

[29] Yildirim O, Talo M, Ay B, Baloglu UB, Aydin G et al. Automated detection of diabetic subject using pre-trained 2D-CNN models with frequency spectrum images extracted from heart rate signals. Comput Biol Med 2019; 113: 103387.

[30] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems , USA; 2012. pp. 1097-1105.

[31] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv 2014; 1409.1556.

[32] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016; 770-778.

[33] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S et al. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition 2015. pp. 1-9.

[34] Chaudhary S, Taran S, Bajaj V, Sengur A. Convolutional Neural Network Based Approach Towards Motor Imagery Tasks EEG Signals Classification. IEEE Sens 2019; 19 (12): 4494–4500.

[35] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016; 770-778.

[36] Brodzicki A, Piekarski M, Kucharski D, Jaworek-Korjakowska J, Gorgon M. Transfer Learning Methods as a New Approach in Computer Vision Tasks with Small Datasets. Foundations of Computing and Decision Sciences 2020; 45 (3): 179-193.

[37] Fawaz IH, Forestier G, Weber J, Idoumghar L, Muller PA. Deep learning for time series classification: a review. Data Mining and Knowledge Discovery 2019; 33 (4): 917–963.

[38] Garg D, Verma GK. Emotion recognition in valence-arousal space from multi-channel EEG data and wavelet based deep learning framework. Procedia Computer Science 2020; 171: 857-867.

[39] Silva A, Zarzo A, González JMM, Munoz-Guijosa JM. Early fault detection of single-point rub in gas turbines with accelerometers on the casing based on continuous wavelet transform. Journal of Sound and Vibration 2020; 487: 115628.

[40] Li K, Su L, Wu J, Wang H, Chen P. "A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine, Applied Sciences 2017; 7 (10).

[41] Ju B, Zhang H, Liu Y, Liu F, Lu S et al. "A feature extraction method using improved multi-scale entropy for rolling bearing fault diagnosis," Entropy 2018; 20 (4): 212.

[42] Ge J, Yin G, Wang Y, Xu D,Wei F. "Rolling-Bearing Fault-Diagnosis Method Based on Multimeasurement Hybrid-Feature Evaluation," Information 2019; 10 (11): 359.

[43] Zhu K, Chen L,Hu X. "A Multi-scale Fuzzy Measure Entropy and Infinite Feature Selection Based Approach for Rolling Bearing Fault Diagnosis," Journal of Nondestructive Evaluation 2019; 38 (4): 90.