Gearbox Multiple Faults Diagnosis under Stationary and Non-Stationary Operating Conditions Using Convolutional Neural Networks

Destine Mashava[1,*], James Kuria Kimotho[2], Onesmus Mutuku Muvengei[2]

[1]Department of Mechanical Engineering, Pan African University Institute for Basic Sciences, Innovation and Technology, Nairobi, Kenya.
[2]Department of Mechanical Engineering, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya.

Abstract

High accuracy in gearbox fault diagnosis is of paramount importance for keeping industrial systems safe and working normally. Concerning various single or multiple faults diagnosis using classical machine learning algorithms, the feature extraction and selection methods are time-consuming and labor-intensive processes requiring expert knowledge of the relevant features related to the system. To mitigate this problem, a deep learning convolutional neural network (CNN) is proposed which enables automatic feature learning from the time-frequency domain image representations data input. The proposed approach employs a CNN model which uses max-pooling and batch normalization between each convolution for training acceleration and reduction of generalization error. The proposed model performance in multiple faults diagnosis is evaluated using gearbox vibration data obtained under stationary operating conditions. The proposed model classification performance is also evaluated using engineered non-stationary operating vibrations data sets. To achieve this, data sets were engineered using gearbox stationary operating conditions vibration data obtained using five operating speeds (30 Hz, 35 Hz, 40 Hz, 45 Hz, and 50 Hz) under low and high load conditions. The non-stationary operating conditions data sets were developed based on three operational conditions namely; constant operating speed and variable load, variable operating speed and constant load, and variable operating speed and variable load. This methodology was applied to PHM2009 gearbox vibration data sets which consist of multiple component faults. Time domain, frequency domain, and time-frequency analysis methods (short-time Fourier transform (STFT), continuous wavelet transform (CWT), empirical mode decomposition (EMD), Wigner Ville distribution (WVD), and wavelet synchrosqeezed transform (WSST)), were investigated for their effectiveness in multiple faults diagnosis. The proposed CNN architecture exhibited high classification performance as high as 99.9% under stationary operating conditions in multiple faults diagnosis as compared to other architectures. The use of image input to the CNN model gave better performance compared to feature vector input. The application of the developed diagnostic model for multiple fault diagnosis under non-stationary operating conditions gave satisfactory classification performance as high as 86.5% using scalogram images. This work provides the possibility that stationary operating conditions-based diagnostic models, can be deployed for multiple faults diagnosis on industrial equipment operating under non-stationary conditions.

Keywords

1. Introduction

A gearbox, being an important component in power transmission for many industrial machines and can have faults mostly due to fatigue cracking under cyclic contact stress and wear [1, 2]. Gearbox fault diagnosis which is presently utilized in the industry is largely dependent on the vibration signals which are collected from the operational gearbox and to maintain overall system availability, the gearbox requires condition monitoring to avoid equipment sudden failure and secondary damage [3, 4]. Gearbox component defects can include bearing spalling, gear teeth breakage, chipped teeth, surface damage, and tooth cracks which after some time will create some form of vibration that will fail supporting components such as the bearings, and shafts [4]. To enable real-time gearbox condition monitoring, machine learning concepts can be implemented on the gearbox while it is in service and this has a wide range of economic benefits to the industry through opening new streams of revenue by allowing original equipment manufacturers to continuously contribute to the availability and maintainability of equipment [5]. However, the drawback with the current machine learning algorithms is the tedious process of developing a feature vector from the sensor-measured raw signal by feature extraction [6, 7, 8]. Hence, after extraction of features, machine learning methods such as Support Vector Machine, Random forest, or K-Nearest Neighbour are then used to separate, classify, and predict faults from the learned patterns present within the layers of the feature vector [9, 10]. The fact that the feature vectors are human-engineered, makes them more susceptible to uncertainty and bias in the domain of the engineers creating feature vectors. It also has been noted that as the data size becomes substantial, feature vector development becomes complex and time-consuming, resulting in the need for an automatic feature extraction approach to mitigate the prior mentioned drawbacks of manual feature extraction [5].

Therefore, regarding the prior mentioned drawbacks of feature engineering in machine learning algorithms, deep learning is brought into consideration due to its ability to represent learning methods consisting of multiple layers [5, 11]. The main advantage of the deep learning approach is its ability to learn nonlinear depictions of raw signals to a higher level of abstraction and complexity without a human engineer's intervention in the learning process [12]. The use of deep learning using convolutional neural networks (CNNs) has become dominant in handling the complexity of image data input classification. The algorithms have become dominant to outperform human accuracies for the same tasks in fields such as analysis of medical images, computer vision, etc. [12, 13, 14]. This ability of deep learning models to automatically extract features from raw input data such as vibration signals has brought valuable opportunities to the fault diagnosis and the engineering field [5]. Deep learning methodologies have been recently implemented in fault diagnostics of rolling element and gear signals [11, 14, 15, 16]. Wang et al. [7] also proposed the use of wavelet transform images as input to a convolutional neural network for bearing fault detection within a set of vibration data. Although they achieved classification accuracy as high as 95%, their work was restricted to non-time varying operating conditions and only wavelet transform input data was considered in the diagnosis of the rotor system.

Wang et al. [15] proposed a convolutional neural network model for motor fault diagnosis, in which a short-time Fourier transform was implemented in raw signal pretreatment to acquire time-frequency inputs. Their model achieved testing accuracy of up to 100%, but their work was only based on short-time Fourier transform and motor faults. There remains an opportunity to explore the performance of other time-frequency domain representations and the effect of time-varying conditions on the diagnosis model. Guo et al. used Case Western's bearing vibration data set [17] in their research work to diagnose faults and severity using an adaptive deep convolutional neural network (ADCNN). They managed to achieve high classification performances as high as 99% using raw vibration signals obtained under stationary operating conditions. However, their work can be further explored by taking into consideration the effects of multiple component faults and time-varying conditions in the development of the ADCNN model.

Chen et al. [11] proposed the use of a feature vector as input to a convolutional neural network architecture consisting of one convolutional layer and one pooling layer in diagnosing gearbox bearing and gear faults using vibration data. They managed to achieve model classification performances as high as 98% based on time and frequency domain feature vector input. There remains an opportunity to explore time-frequency domain representations, multiple hybrid faults, and time-varying conditions in diagnostic model development. Zhou et al. [18] implemented fast Fourier transform on Case Western Reserve University bearing vibration data set to generate a feature vector input for a deep neural network (DNN) consisting of three, four, and five hidden layers, as the deep neural networks do not consist of convolutional and pooling layers, but only consist of hidden layers. In their work high model performances as high as 99.9% were obtained, but their work was limited to bearing faults, frequency domain analysis, and non-time varying operating conditions. Liu et al. [14] used spectrograms as input vectors for sparse and stacked autoencoders consisting of two hidden layers. The results obtained by the researchers indicated difficulty in classifying outer race faults versus the baseline. However, the work can be further explored using other time-frequency domain representations such as continuous wavelet transform scalogram images in the development of the diagnostic model.

The reviewed deep learning models and their applications to fault diagnosis are usually limited by their sensitivity to

experimental noise or their reliance on traditional feature extraction [5]. Fault diagnosis research work is also mainly centered on determining a single fault type such as bearing fault or gear fault. The previous models were also developed based on stationary operating conditions vibration data, which is not a true representation of industrial equipment operating conditions such as a gearbox in which its vibration data signature is non-stationary due to variations in its rotational speed and the applied load.

The study aims to propose a deep learning convolutional neural network (CNN) architecture based on time-frequency image analysis for multiple fault diagnosis of a two-stage industrial gearbox under stationary and non-stationary operating conditions. The main model aspect consists of convolution layers with a max-pooling layer and a batch normalization layer, between each convolutional layer to speed up the network training and classification performance. The CNN architecture consists of five convolutional layers established based on a maximum image input size of 128×128 pixels. The paper also introduces the development and engineering of non-stationary operating conditions vibration data sets from stationary operating conditions vibration data. The engineered vibration data sets are used to evaluate the developed CNN diagnosis model classification performance in multiple faults diagnosis of a gearbox operating under non-stationary conditions. Five time-frequency methods (continuous wavelet transform (CWT) scalogram, Short-Time Fourier transform (STFT) spectrogram, wavelet synchrosqeezed transform (WSST), Wigner Ville Distribution (WVD), empirical mode decomposition (EMD) in which the image is processed using the Hilbert-Huang transformation (HHT)) are used as data input to the model and their effectiveness in multiple faults diagnosis is also evaluated. These five methods are utilized to enable suitable outputs for the discovery of complex and high-dimensional representations without the need for additional feature extraction. The deep learning model performance in multiple fault diagnosis is also evaluated based on statistical feature vector input.

In addition to the convolutional neural network architecture and five time-frequency analysis methods and statistical feature engineering, this paper also examines the proposed model performance based on variable image size input of 128×128, 96×96, 64×64, and 32×32 pixels. It is observed that varying the input image size has a significant impact on the convolutional neural network's number of learnable parameters. The training time is observed to decrease with a reduction in input image size, while the classification accuracy is negatively impacted. The proposed model is evaluated based on the PHM2009 gearbox vibration data sets which consist of eight health conditions for spur gear type and six health conditions for helical gear-type [19].

The rest of the paper is organized as follows: Section 2 outlines the fault diagnosis approach overview and the proposed convolutional neural network (CNN) architecture. Section 3 provides the methodology application to the PHM2009 gearbox vibration data set in multiple fault diagnosis under various signal processing techniques and input data types. Section 4 outlines the engineering of non-stationary operating conditions vibration data sets and implementation of the signal processing time-frequency domain to generate trained model testing data. Section 5 provides model validation under stationary operating conditions. Section 6 concludes with the discussion of the results on the proposed approach in which stationary operating conditions vibration data sets are used to develop a deep learning diagnostic model for multiple faults classification under time-varying operating conditions. The approach managed to give high classification accuracy using stationary operating conditions vibration data sets and when the model was used to diagnose faults under non-stationary operating conditions, satisfactory performance was obtained using scalograms as input. The proposed approach reveals that diagnostic models developed using non-time-varying conditions data can be successfully deployed on industrial equipment running under time-varying conditions. This approach could be of paramount importance to the industry, as it will enable to circumvent the difficulty associated with obtaining model development health pattern data from newly installed equipment.

2. Approach to Gearbox Multiple Faults Diagnosis

The approach applied in the development of the convolutional neural network diagnosis model applied for multiple faults diagnosis under stationary and non-stationary operating conditions of an industrial gearbox is shown in Figure 1. In this approach, gearbox vibration data sets consisting of multiple component fault, patterns are analyzed in the time-frequency domain to generate image representations used as data input in the training, testing, and validation of the proposed deep learning architecture based on the implementation of max-pooling and batch normalization between each convolutional layer. The gearbox vibration data sets obtained under stationary operating conditions are processed in the time and frequency domain to generate statistical feature vectors which are used to evaluate the performance of the deep learning model in multiple faults diagnosis based on the manually engineered feature vector data. To evaluate the developed classification model performance under time-varying operating conditions, gearbox vibration data sets obtained under stationary operating are engineered to create non-stationary operating conditions vibration data sets. The non-stationary operating conditions vibration data sets are processed in the time-frequency domain to generate image representations which are used as input data for the classification model. The developed classification model is then validated using stationary operating conditions time-frequency domain input data to confirm the developed model ef-
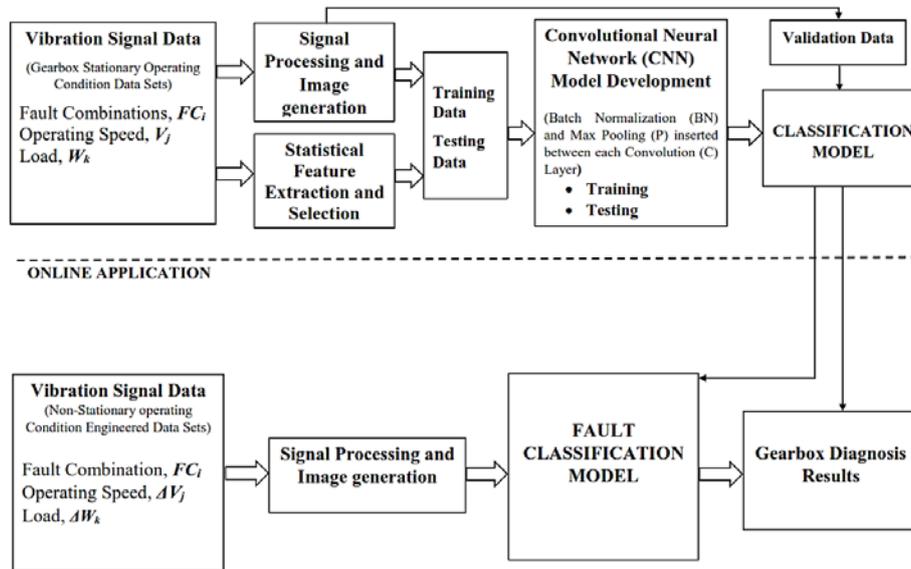
fectiveness in multiple faults diagnosis.

Figure 1. Gearbox multiple fault diagnosis overall approach.

2.1. Gearbox Multiple Faults Diagnosis Proposed Convolutional Neural Network Architecture

The primary element of the proposed architecture consists of a batch normalization layer between convolution layers, the batch normalization layers are essential in the normalization of activations and gradients propagating through the network, allowing to speed up network training and reduction of network initialization sensitivity [20]. The batch normalization layer is then followed by a nonlinear activation function and the rectified linear unit function is used due to its faster training ability. After the nonlinear activation function, a max-pooling layer is inserted between each convolutional layer with an activation function. Max pooling is considered due to its high ability in the extraction of low-level features which cannot be correctly identified by average pooling. This model architecture section executes the automatic feature extraction process and is then followed by a fully connected layer which refines the features which are then passed to the final decision layer, the Softmax layer, which enables effective multiclass classification by assigning decimal probabilities to each class. Figure 2 gives the proposed convolutional neural network architecture applied in gearbox multiple faults diagnosis.
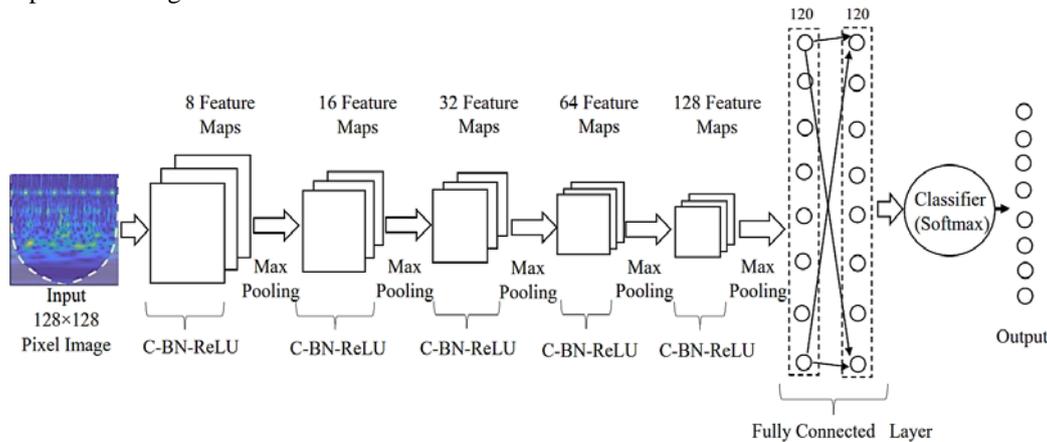
Figure 2. Proposed Convolutional Neural Network Architecture, C-Convolution, BN-Batch Normalization, ReLU- Rectified Linear Unit Function.

The proposed architecture first convolutional layer consists of 8 feature maps of 3×3 size, followed by a batch normalization layer, then a rectified linear unit layer, and lastly is a max-pooling pooling layer of 8 feature maps of 2×2 size. This makes up the first stage. The second stage consists of a convolutional layer of 16 feature maps each of 3×3 size, followed by a batch normalization layer, rectified linear unit layer, and a max-pooling layer of 16 feature maps of

2×2 size. The third stage consists of a convolutional layer of 32 feature maps each, of 3×3 size, followed by a batch normalization layer, rectified linear unit layer, and a max-pooling layer of 32 feature maps of 2×2 size. The fourth stage consists of a convolutional layer of 64 feature maps each, of 3×3 size, followed by a batch normalization layer, rectified linear unit layer, and a max-pooling layer of 64 feature maps of 2×2 size. The fifth stage consists of a convolutional layer of 128 feature maps each, of 3×3 size, followed by a batch normalization layer, rectified linear unit layer, and a max-pooling layer of 128 feature maps of 2×2 size. The sixth layer consists of a fully connected layer of 120 features. The intent of increasing the number of filters at each stage by the power of two is to increase the representational power of the architecture as the initial layers learn primitive regularities in the data, but, going through to the upper layers many obstructions are encountered. Therefore, the need to increase the number of feature maps at each stage to compensate for the obstructions. The increase in the number of feature maps at each stage is done to compensate for the resolution loss of the input image. The model five convolutional layers were established based on an input image size of 128×128 pixels. The proposed model structure constrain to five convolutional layers is also meant to eliminate computational complexity, model over-fitting, and to maintain optimum final layer feature maps which are critical to the overall model performance.

Table 2 provides an overview of the convolutional neural network proposed architecture and the other architectures implemented in multiple fault diagnosis, where C is a convolutional layer, BN is a batch normalization layer, ReLU is Rectified Linear Unit layer, P is a pooling layer, and FC is a fully connected layer. The number preceding the C, P, and FC indicates the number of feature maps used. The dimensions $[3 \times 3]$ and $[2 \times 2]$ indicate the pixel size of the features. Training the CNN involves the learning of all weights and biases present within the architecture. The weights and biases are referred to as learnable parameters. The convolutional networks also require the specification and optimization of the hyper-parameters, learning rate, and dropout. Dropout is applied to the convolutional neural network to prevent over-fitting, reduce training error, and effectively thin the network [21]. For the other hyper-parameters, the learning rate stochastic gradient descent with momentum (SGDM) algorithm is used for learning rate optimization. The processing equipment for the analysis is as follows; CPU Core i7-8565U, 2.1GHz with 16 GB RAM and GPU AMD 6500, and Intel 520.

Table 1. Convolutional Neural Network architectures applied in multiple fault diagnosis

Model	Architecture
Architecture1 Wang et al. [15]	Input Layer[128x128]-64C[3x3]-64P[2x2]-64C[4x4]-64P[2x2]-128C[3x3]- 128P[2x2]-FC[512]
Architecture 2 Verstraet et al.	Input Layer[128x128]-32C[3x3]-32C[3x3]-32P[2x2]-64C[3x3]-64C[3x3]-64P[2x2]- 128C[3x3]-128C[3x3]-128P[2x2]-FC[100]-FC[100]
Architecture 3	Input Layer[128x128]-8(C[3x3]-BN-ReLU)-8P[2x2]- 16(C[3x3]-BN-ReLU)-16P[2x2]-32(C[3x3]-BN-ReLU)-32P[2x2]-FC[120]-FC[120]
Architecture 4	Input Layer[128x128]-8(C[3x3]-BN-ReLU)-8P[2x2]- 16(C[3x3]-BN-ReLU)-16P[2x2]-32(C[3x3]-BN-ReLU)-32P[2x2]- 64(C[3x3]-BN-ReLU)-64P[2x2]-FC[120]-FC[120]
Proposed Architecture	Input Layer[32x32]-8(C[3x3]-BN-ReLU)-8P[2x2]- 16(C[3x3]-BN-ReLU)-16P[2x2]-32(C[3x3]-BN-ReLU)-32P[2x2]- 64(C[3x3]-BN-ReLU)-64P[2x2]-128(C[3x3]-BN-ReLU)-128P[2x2]-FC[120]-FC[120]
	Input Layer[64x64]-8(C[3x3]-BN-ReLU)-8P[2x2]- 16(C[3x3]-BN-ReLU)-16P[2x2]-32(C[3x3]-BN-ReLU)-32P[2x2]- 64(C[3x3]-BN-ReLU)-64P[2x2]-128(C[3x3]-BN-ReLU)-128P[2x2]-FC[120]-FC[120]
	`Input Layer[96x96]-8(C[3x3]-BN-ReLU)-8P[2x2]- 16(C[3x3]-BN-ReLU)-16P[2x2]-32(C[3x3]-BN-ReLU)-32P[2x2]- 64(C[3x3]-BN-ReLU)-64P[2x2]-128(C[3x3]-BN-ReLU)-128P[2x2]-FC[120]-FC[120]
	Input Layer[128x128]-8(C[3x3]-BN-ReLU)-8P[2x2]- 16(C[3x3]-BN-ReLU)-16P[2x2]-32(C[3x3]-BN-ReLU)-32P[2x2]- 64(C[3x3]-BN-ReLU)-64P[2x2]-128(C[3x3]-BN-ReLU)-128P[2x2]-FC[120]-FC[120]
Architecture 5	FeatureInputLayer [6x1]-FC[120]-BN[120]-ReLU[120]-FC[6]
Architecture 6	FeatureInputLayer [8x1]-FC[120]-BN[120]-ReLU[120]-FC[8]

3. Data Processing and Architecture Training

The applied PHM2009 gearbox vibration data sets consisted of 266656 data points per gearbox health condition sam-

ple [19]. The gearbox spur gear type vibration data set contains eight different health conditions and the helical gear-type includes six different health conditions. The vibration data sets consist of multiple faults for gears, shafts, and bearings. The total images produced from the vibration data sets for spur gear type were 10240 images and 7680 images for helical gear per each time-frequency domain analysis (CWT, EMD-HHT, STFT, WSST, and WVD) method. The generated images exhibit unique pixel patterns and intensities per each gearbox health case and this enables the CNN model to accurately classify the gearbox health condition classes. The generated images were then resized into 32×32, 64×64, 96×96, and 128×128-pixel image sizes using MATLAB functions. The resizing is done to determine the most suitable image size for training and testing the proposed CNN model architecture. Based on size and quantity, the 128×128, 96×96, 64×64, and 32×32 images for helical gear-type result in 125 829 120, 70 778 880, 31 457 280, and 7 864 320 pixel data points, and for spur gear type 167 772 160, 94 371 840, 41 943 040 and 10 485 765 pixel data points respectively which are used by the proposed architecture to learn the data features. From the image data, a sixty /twenty/twenty percentage split is used for the training, testing, and validation set, respectively. The prediction accuracies of the architectures given in Table 1 are evaluated using 128×128 pixel images. The proposed model architecture is also trained and tested under the four image sizes, of 32×32, 64×64, 96×96, and 128×128 pixels to determine the effect of varying the input image resolution on the prediction accuracy in multiple faults diagnosis. The generated 128×128-pixel time-frequency domain image representations for the gearbox baseline vibration response are shown in Tables 2 and 3, for helical gear and spur gear type respectively.

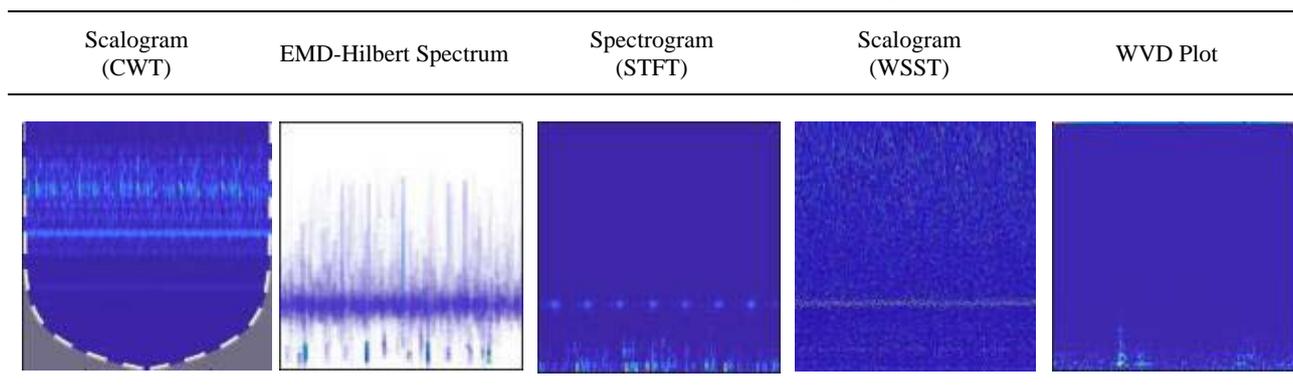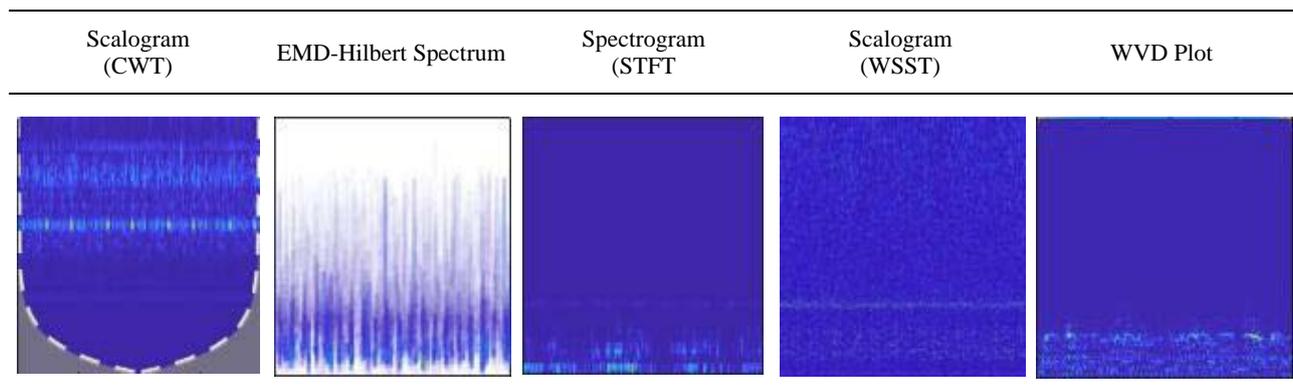**Table 2. Gearbox baseline vibration response 128×128 pixel images for helical gears**

Scalogram (CWT)	EMD-Hilbert Spectrum	Spectrogram (STFT)	Scalogram (WSST)	WVD Plot

Table 3. Gearbox baseline vibration response 128×128 pixel images for spur gears

Scalogram (CWT)	EMD-Hilbert Spectrum	Spectrogram (STFT	Scalogram (WSST)	WVD Plot

The first f convolutional neural network architectures given in Table 1, are trained using 128×128-pixel images. The model architectures were trained using sixty percent of the data which corresponds to 760 images per health pattern and twenty percent of the data which corresponds to 256 images per health pattern was used in model testing. The classification performance measures for the architectures are given in Tables 4 and 5 for helical and spur gear types respectively. The overall multiclass classifier accuracy is determined as the average accuracy per class and is essential in determining the model's ability to correctly classify true positive and true negative data. Assessment of precision is essential in diagnostic systems as it emphasized false positives, and this enables the evaluation of the model's ability to predict actual faults. Model precision is determined from the average precision of all health patterns used in model training. For the PHM2009 vibration data, six classes were used under helical gear data and eight classes were used under the spur gear data in architectures training.

Tables 4 and 5, outline the mean accuracy, average precision, average sensitivity, average specificity, and average F-score of the classes for the three architectures under helical and spur gear data input data respectively. Recall or sensitivity is another effective performance measure for a diagnostic system as it evaluates the model's ability in predicting real faults. Hence, sensitivity emphasized false negatives. Specificity, or true negative rate, emphasizes false positives and is effective for examining false alarm rates. The results obtained from the developed model show high specificity and sensitivity implying that the model has a higher ability to accurately diagnose faults with minimum false alarms. The F-score metric assesses the harmonic balance between precision and sensitivity. Although it does not consider true negatives, it illustrates a diagnostic system's ability to accurately predict true faults. Overall, the proposed architecture outperforms the other architectures in multiple faults diagnosis of the industrial gearbox using time-frequency domain image representations as data input as shown by performance metrics in Tables 4 and 5 under the five time-frequency domain signal processing techniques (CWT, EMD, STFT, WSST, and WVD) data input.

Table 4. Models performance measures based on 128×128 pixel images input for helical gear type

Input Image (128x128)	Model	Model Performance Metrics				
		Accuracy	Precision	Sensitivity	Specificity	F-Score
CWT	Architecture 1	98.57%	98.57%	98.66%	99.72%	98.57%
	Architecture 2	99.41%	99.41%	99.42%	99.88%	99.41%
	Architecture 3	98.89%	98.91%	98.89%	99.78%	98.89%
	Architecture 4	99.21%	99.24%	99.22%	99.84%	99.22%
	Proposed Model	99.67%	99.67%	99.68%	99.94%	99.67%
EMD	Architecture 1	91.88%	91.08%	91.16%	98.27%	91.88%
	Architecture 2	93.62%	93.62%	94.29%	98.75%	93.62%
	Architecture 3	92.32%	93.31%	92.32%	98.46%	92.32%
	Architecture 4	93.81%	93.96%	93.82%	98.76%	93.78%
	Proposed Model	98.11%	98.11%	98.16%	99.62%	98.11%
STFT	Architecture 1	79.70%	79.7%	79.99%	95.94%	79.70%
	Architecture 2	89.19%	89.19%	89.42%	97.85%	89.17%
	Architecture 3	83.66%	86.70%	83.66%	96.73%	83.20%
	Architecture 4	87.56%	87.58%	87.57%	97.51%	87.47%
	Proposed Model	96.94%	96.94%	96.98%	99.39%	96.94%
WSST	Architecture 1	87.70%	87.70%	87.59%	97.57%	87.70%
	Architecture 2	96.55%	96.55%	96.58%	99.31%	96.55%
	Architecture 3	96.09%	96.09%	99.09%	99.22%	96.07%
	Architecture 4	95.38%	95.38%	95.38%	99.08%	95.37%
	Proposed Model	98.37%	98.37%	98.40%	99.68%	98.37%
WVD	Architecture 1	85.03%	85.03%	85.13%	97.01%	85.02%
	Architecture 2	91.08%	91.08%	91.90%	98.24%	91.15%
	Architecture 3	88.35%	90.69%	88.35%	97.67%	88.32%
	Architecture 4	91.34%	92.05%	91.34%	98.27%	91.26%
	Proposed Model	98.05%	98.05%	98.16%	99.62%	98.05%

Table 5. Models performance measures based on 128×28 pixel images input for spur gear type

Image Type (128x128)	Model	Model Performance Metrics				
		Accuracy	Precision	Sensitivity	Specificity	F-Score
CWT	Architecture 1	98.93%	98.93%	99.96%	99.95%	98.93%
	Architecture 2	99.46%	99.46%	99.47%	99.92%	99.46%
	Architecture 3	99.76%	99.76%	99.76%	99.97%	99.76%
	Architecure 4	99.56%	99.56%	99.56%	99.94%	99.56%
	Proposed Model	99.90%	99.90%	99.90%	99.99%	99.90%
EMD	Architecture 1	97.17%	98.17%	97.24%	99.60%	97.18%
	Architecture 2	98.34%	98.34%	99.35%	99.76%	99.34%
	Architecture 3	98.00%	98.15%	99.00%	99.71%	97.97%
	Architecure 4	98.24%	98.25%	98.24%	99.75%	98.23%
	Proposed Model	99.61%	99.61%	99.61%	99.94%	99.61%
STFT	Architecture 1	97.66%	97.66%	97.66%	99.67%	97.70%
	Architecture 2	98.93%	98.93%	98.96%	99.85%	98.93%
	Architecture 3	98.39%	98.40%	98.39%	99.77%	98.38%
	Architecure 4	98.56%	98.59%	98.57%	99.71%	98.57%
	Proposed Model	99.66%	99.66%	99.66%	99.95%	99.66%
WSST	Architecture 1	99.41%	99.41%	99.42%	99.92%	99.41%
	Architecture 2	99.61%	99.62%	99.62%	99.94%	99.61%
	Architecture 3	99.46%	99.47%	99.46%	99.92%	99.46%
	Architecure 4	99.22%	99.24%	99.24%	99.89%	99.22%
	Proposed Model	99.85%	99.85%	99.86%	99.98%	99.85%
WVD	Architecture 1	98.01%	98.01%	98.02%	99.72%	98.01%
	Architecture 2	99.02%	99.02%	99.02%	99.86%	99.02%
	Architecture 3	98.44%	98.48%	98.44%	99.78%	98.43%
	Architecure 4	99.17%	99.19%	99.17%	99.88%	99.77%
	Proposed Model	99.56%	99.56%	99.58%	99.94%	99.56%

The metric parameters given in Tables 4 and 5, which are accuracy, precision, sensitivity, specificity, and F-Score are obtained from a trained model confusion matrix results. Figure 3 shows the proposed architecture confusion matrices for 128 × 128 pixels scalogram input images under helical and spur gear types. The confusion matrix under helical gear-type gives an average model testing accuracy of 99.7% for the 6 health conditions cases in which only five image data samples were wrongly predicted to belong to a healthy class HAHS but actually, they belong to fault class HGFS which represent shaft faults. The confusion matrix under spur gear type gives an average model testing accuracy of 99.90% for the 8 health conditions cases in which only two image data samples were wrongly predicted to belong to faulty class SFFS, but actually, they belong to fault class SGFS. The scalogram confusion matrix results shown in Figure 3, confirm that the developed multiple faults diagnosis model can be effectively applied for machinery fault diagnosis with minimal false alarm rate.
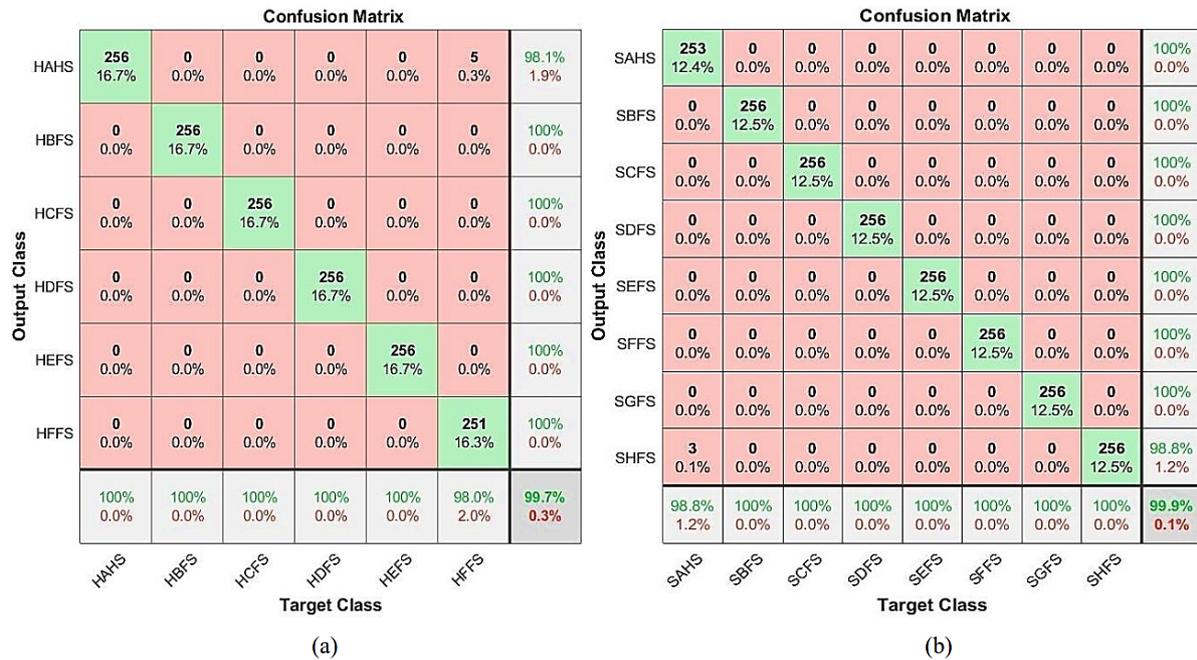
Figure 3. Confusion Matrix for the Proposed Architecture (a) helical gear type (b) spur gear type.

The proposed model architecture multiple faults classification performance measurements based on variable image sizes of (128 × 128, 96 × 96, 64 × 64, and 32 × 32) pixels are given in Tables 6 and 7, for helical gear type and spur gear type respectively. The loss of resolution due to the reduction in image size from 128 × 128 to 96 × 96 to 64 × 64 to 32 × 32 pixels influenced the fault diagnosis in which there was a slight drop in the model accuracy as the input image sizes were reduced. The scalograms were least affected by resolution reduction compared to the spectrograms and Hilbert spectrum plots for helical gears input data. Also under spur gears input, the changes in input image resolution had the least effect on scalograms compared to the other time-frequency domain image inputs, which are spectrograms, WVD plots, and Hilbert Spectrum plots.

Table 6. Proposed architecture accuracies based on variable image sizes for helical gear type

Data Input	32 × 32 pixel images	64 × 64 pixel images	96 × 96 pixel Image	128 × 128 pixel Images
CWT-Scalogram	98.96%	99.35%	99.41%	99.61%
EMD-Hilbert Spectrum	91.80%	94.14%	96.09%	98.11%
STFT-Spectrogram	90.17%	93.74%	94.99%	96.94%
WSST-Scalogram	93.88%	96.48%	97.46%	98.31%
WVD Plots	92.48%	94.01%	94.92%	98.05%

Table 7. Proposed architecture accuracies based on variable image sizes for spur gear type

Data Input	32 × 32 pixel images	64 × 64 pixel images	96 × 96 pixel image	128 × 128 pixel images
CWT-Scalogram	99.61%	99.80%	99.85%	99.90%
EMD-Hilbert Spectrum	97.75%	98.63%	98.93%	99.61%
STFT-Spectrogram	98.62%	99.32%	99.41%	99.85%
WSST-Scalogram	98.88%	99.41%	99.51%	99.90%
WVD Plots	98.00%	98.83%	99.02%	99.56%

Table 8 gives the model accuracy and training time for CWT scalogram variable-sized images for helical and spur gear types. The model training time had a significant decrease with a slight reduction in accuracy as the input image resolution was being reduced as shown in Table 8. Therefore, this observation can be utilized to minimize computation-

al costs and to achieve rapid model deployment.

Table 8. Proposed architecture accuracies and training time for CWT scalogram input images

Input Data Type	Gear Type	Image Size (pixels)	Accuracy	Training Time (min)
Scalogram (CWT)	Helical	32 × 32	98.96%	1.75
		64 × 64	99.35%	5.10
		96 × 96	99.41%	11.35
		128 × 128	99.61%	18.00
	Spur	32 × 32	99.61%	2.35
		64 × 64	99.80%	7.80
		96 × 96	99.85%	16.70
		128 × 128	99.90%	27.80

3.1 Statistical Features Extraction and Architecture Performance Evaluation

To have a comparison between the automatic feature extraction approach and the manual feature extraction approach, statistical feature vectors are generated from the gearbox vibration data and are used as input for training and testing the deep learning model. The vibration signals are divided into bins of 8333 samples, each with a 50% overlap. Each of these bins is further processed to extract features from the original signal under the time and frequency domain. mean, kurtosis, shape factor, standard deviation, and root mean square resulted to be the best features selected using MATLAB one-way ANOVA function under time-domain analysis. The other three features of the feature vector are the Fast Fourier Transform (FFT) root mean square, the gearbox operating speed, and applied load. The resultant feature vector consisted of 7680 × 8 features for helical gear, while the spur gear type consisted of 10240 × 8 features. Table 9 and Table 10 show the generated statistical feature vectors under helical and spur gear type.

Table 9. Feature vector illustration for helical gear-type vibration data

FFTRMS	Kurtosis	Shape Factor	Standard Deviation	Mean	RMS	Speed	Load	Label
0.02182	5.2835	1.0118	0.0033224	0.021566	0.021807	1	1	HAHS
0.021785	4.2454	1.0116	0.0032896	0.021535	0.021807	1	1	HAHS
0.021799	5.1955	1.0129	0.0034742	0.02152	0.022299	1	1	HAHS
0.021805	5.0621	1.0125	0.0034095	0.021536	0.022297	1	1	HAHS
0.021801	3.9897	1.0116	0.0032903	0.021551	0.021902	1	1	HAHS
0.021833	5.3143	1.0121	0.003365	0.021573	0.02189	1	1	HAHS
0.021792	3.9724	1.0111	0.0032212	0.021552	0.022715	1	1	HAHS
0.021827	5.8413	1.0134	0.0035463	0.021537	0.02272	1	1	HAHS
0.021838	5.9256	1.0135	0.0035577	0.021546	0.022068	1	1	HAHS
0.021778	4.4108	1.0114	0.0032588	0.021533	0.022065	1	1	HAHS
0.021835	5.6332	1.013	0.0034887	0.021554	0.022068	1	1	HAHS
0.02177	4.0349	1.0112	0.003237	0.021528	0.022879	1	1	HAHS
0.02181	5.4681	1.0133	0.0035248	0.021523	0.022216	1	1	HAHS

Table 10. Feature vector illustration for spur gear type vibration data

FFTRMS	Kurtosis	Shape Factor	Standard Deviation	Mean	RMS	Speed	Load	Label
0.022079	4.373	1.0143	0.0036971	0.021768	0.023721	1	1	SAHS
0.022064	4.6492	1.0142	0.003675	0.021756	0.023936	1	1	SAHS
0.022053	4.7696	1.0134	0.0035742	0.021761	0.022629	1	1	SAHS
0.022044	4.3939	1.0137	0.0036174	0.021746	0.022587	1	1	SAHS
0.022067	4.918	1.0151	0.0037879	0.021739	0.24069	1	1	SAHS
0.022059	4.5237	1.0137	0.0036095	0.021761	0.024142	1	1	SAHS
0.022051	4.9422	1.0138	0.0036264	0.02175	0.022012	1	1	SAHS
0.022052	4.5167	1.0141	0.0036699	0.021744	0.022003	1	1	SAHS
0.022045	4.2594	1.0139	0.0036356	0.021743	0.022413	1	1	SAHS
0.022033	4.5153	1.0141	0.003662	0.021727	0.022373	1	1	SAHS
0.02205	4.2795	1.0135	0.0035859	0.021756	0.022071	1	1	SAHS
0.022076	4.8694	1.0145	0.0037252	0.021759	0.022077	1	1	SAHS
0.022066	4.6118	1.0147	0.0037391	0.021747	0.022731	1	1	SAHS

Training and testing of the deep learning model using hand-crafted features was achieved by considering only the feature input layer, batch normalization, and Rectified Linear Unit function while the convolutional layer is eliminated from the architecture as given in Table 1. The architecture performance measures obtained under statistical feature vector input are given in Table 11, for helical and spur gear types.

Table 11. Architecture 3 and 4 performance measures under statistical feature vector data input

Model	Feature Vector	Accuracy	Precision	Sensitivity	Specificity	F-Score
Architecture 5	7680 × 8	82.91%	83.91%	84.52%	96.84%	83.63%
Architecture 6	10240 × 8	82.69%	83.28%	83.55%	97.64%	83.00%

The use of manually engineered features in fault diagnosis was outperformed by the proposed convolutional neural network model based on time-frequency domain input images and also architectures 3 and 4 were outperformed by the shallow convolutional neural network architectures whose performance measures are shown in Tables 4 and 5, signifying a remarkable ability of deep learning architectures in feature extraction and classification against manually engineered feature vectors which are more susceptible to data loss and bias.

4. Non-Stationary Operating Conditions Vibration Data Sets Engineering

The need to determine the feasibility of using a diagnostic model developed based on stationary operating conditions vibration data for fault diagnosis under non-stationary operating conditions necessitated the development of an approach to engineer gearbox vibration data sets obtained under stationary operating conditions into non-stationary operating conditions vibration data sets. This approach was also necessitated by the scarcity of gearbox non-stationary operating conditions vibration data sets. Therefore, due to the challenges faced in obtaining non-stationary operating conditions data from equipment for diagnostic model development, the use of stationary operating conditions gearbox vibration data to engineer non-stationary operating conditions data sets can bring significant benefit in the development of fault diagnosis models for industrial equipment which usually operate under non-stationary conditions.

The non-stationary operating conditions vibration data sets are engineered using PHM2009 gearbox vibration data, which was obtained under stationary operating conditions. Three categories of non-stationary operating conditions vibration data sets are developed based on constant operating speed and variable load, variable operating speed and constant load, and variable operating speed and variable load. The vibration data samples are split into bins and the bins are arranged according to each category pattern to come up with data samples for non-stationary operating conditions. Table 12, gives the bin arrangement pattern per each category of gearbox non-stationary operating conditions.

Table 12. Non-stationary operating conditions data set engineering

Original Data Sample Patterns (Gear Type -Helical-H and Spur-S)			Non-Stationary Operating Conditions Data set Engineering	
Runs (R) Pattern	Operating Frequency (F)	Load (L)	Non-Stationary Conditions	Bins Pattern Per Data Sample
RL1-1	F1-30Hz	L1-High	Variable Load and Constant Operating Speed	[F1L1-F1L2-F1L2-F1L1-F1L2] [F2L1-F2L2-F2L1-F2L2-F2L1] [F3L2-F3L1-F3L2-F3L1-F3L2] [F4L1-F4L2-F4L1-F4L2-F4L1] [F5L2-F5L1-F5L2-F5L1-F5L2]
RL1-2	F2-35Hz	L2-Low	Variable Operating Speed and Variable Load	[F1L1-F2L2-F3L1-F4L2-F5L1] [F2L2-F3L1-F4L2-F5L1,F1L2] [F3L1-F4L2-F5L1-F1L2-F2L1] [F4L2-F5L1-F1L2-F2L2-F3L1] [F5L1-F1L2-F2L3-F3L1-F4L2]
RL2-1	F3-40Hz		Variable Operating speed and Constant Load	[F1L1-F2L1-F3L1-F4L1-F5L1] [F1L2-F2L2-F3L2-F4L2-F5L2]
RL2-2	F4-45Hz			
	F5-50Hz			

The development of non-stationary operating conditions vibration data sets under constant operating speed and variable load is achieved as follows**:** The PHM2009 gearbox data set consisted of two data samples per each load type (high or low) under each operating speed (30 Hz, 35 Hz, 40 Hz, 45 Hz, and 50 Hz). The data sample obtained per each operational condition consists of 266656 points which are split into bins of 5128 points, the bins are then combined to form a data set of 266656 points in such a manner that the data bins are obtained under the same operating speed but different load type are arranged one after the other and this pattern arrangement of bins is repeated until all the data bins are arranged into data samples of 266656 points. The time-domain representation of the engineered non-stationary operating conditions data sample for helical gear-type data is shown in Figure 4a, and Figure 4b and Figure 4c shows the time domain representations of stationary operating conditions data samples used in the engineering of non-stationary operating conditions data samples under constant operating speed and variable loading data engineering pattern.
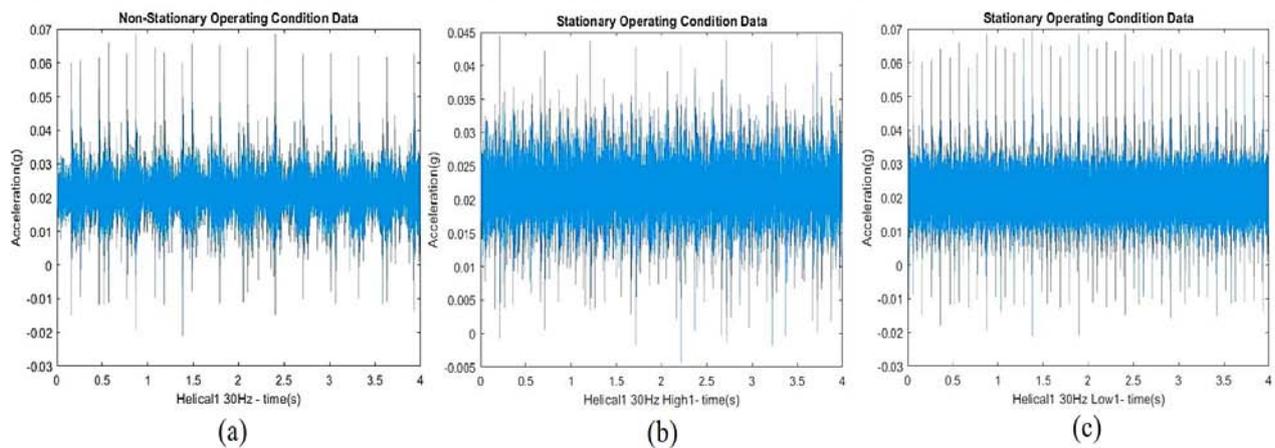

Figure 4. Time-domain representations of (a) Non-stationary operating conditions engineered vibration data sample and stationary operating conditions vibration data samples obtained under (b) low loading at 30Hz operating speed, and (c) high loading at 30Hz operating speed.

The development of non-stationary operating conditions vibration data sets under variable operating speed and constant load is achieved as follows: The data sets obtained under each operating speed (30 Hz, 35 Hz, 40 Hz, 45 Hz, and 50 Hz) are split into bins of 5128 points and the bins from the five sampling speeds are arranged one after the other in such a manner that a bin obtained under 30Hz is followed by a 35Hz bin until to 50Hz bin as per load type (high and low). The bin arrangement pattern is repeated until each data sample has 266656 data points. Figure 5a shows the time-domain representation of the engineered non-stationary operating conditions data sample for baseline spur gear vibration data. The time-domain representations for the stationary operating conditions data samples are shown in Figs.

5b, c, d, e, and f, which are used in the development of non-stationary operating conditions vibration data samples.
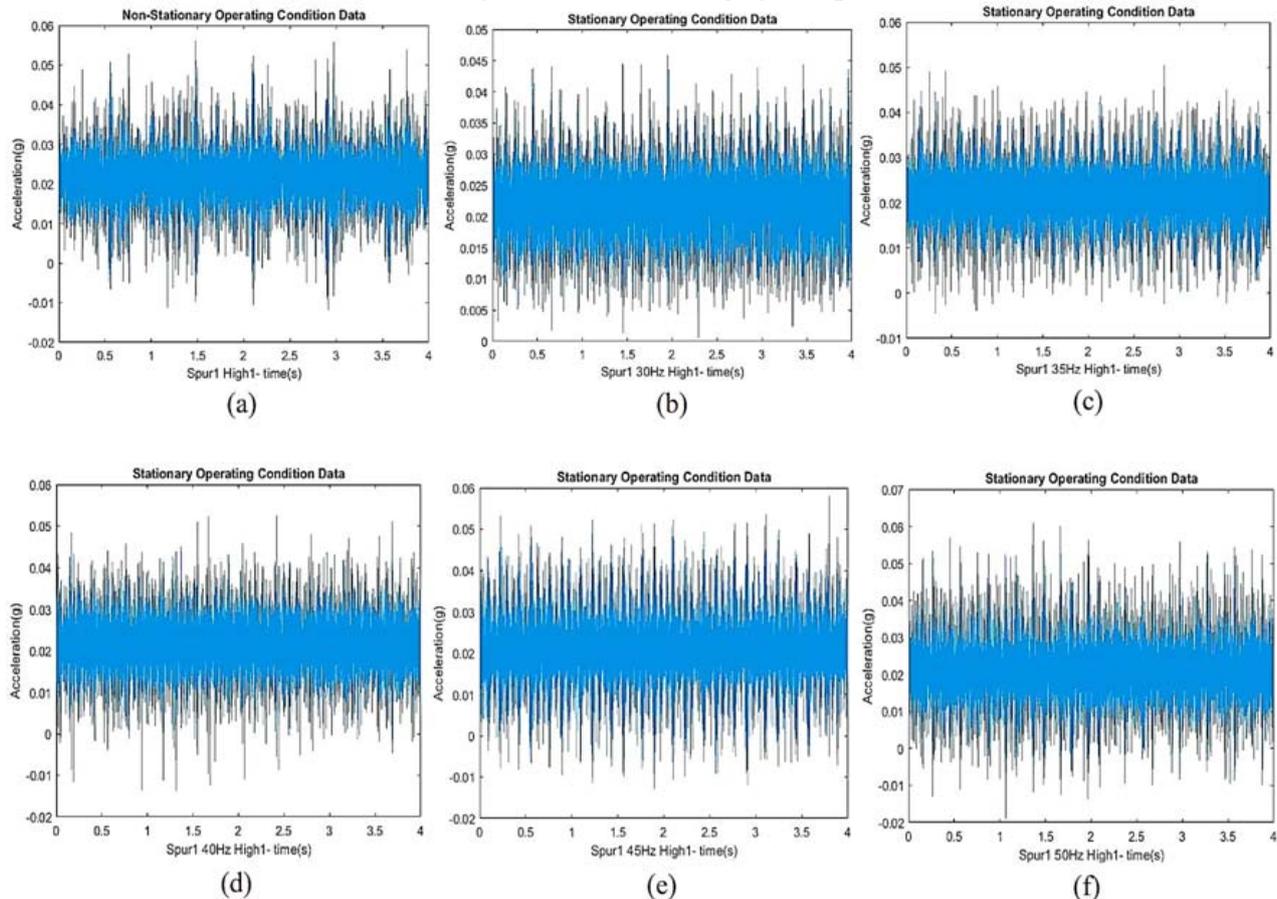

Figure 5. Time-domain representations of (a) Non-stationary operating conditions engineered vibration data sample, and stationary operating conditions vibration data samples obtained at an operating speed of (b) 30Hz and, (c) 35Hz, (d) 40Hz, (e) 45Hz, (f) 50Hz under high loading conditions.

Development of non-stationary operating conditions vibration data sets under variable operating speed and variable load is achieved as follows: The data sets obtained under each operating speed (30 Hz, 35 Hz, 40 Hz, 45 Hz, and 50 Hz) and load type i.e. high and low load are split into bins of 5128 points and the bins from the five operating speeds are arranged one after the other in a manner that if the prior bin was obtained under a lower speed (F1/30Hz at low load (L2) type, the preceding bin sequence will then be a bin obtained under higher operating speed (F2/35Hz) and high load (L1) type as illustrated in Table 12 and this sequence is repeated until each data sample has 266656 data points and the bins pattern arrangement is achieved using MATLAB software.

Figure 6a shows the time-domain representation of the engineered non-stationary operating conditions data sample for baseline helical gear vibration data. The time-domain representations of the five stationary operating conditions vibration data samples used in the engineering of non-stationary operating conditions vibration data samples under variable loading and operating speed are shown in Figs. 6b, c, d, e, and f.

4.1. Non-Stationary Operating Conditions Images Data Generation

The engineered non-stationary operating condition vibration data sets are processed under the time-frequency domain to obtain the input data for testing the proposed convolutional neural network model. The data sets are processed under time-frequency domain using continuous wavelet transform (CWT) scalogram, Hilbert Hung transform (HHT) of intrinsic mode function (IMF) obtained from empirical mode decomposition (EMD), short-time Fourier transform (STFT) spectrograms, and wavelet synchrosqueezed transform (WSST) scalogram and Wigner Ville distribution (WVD) plots. The 128×128 pixels continuous wavelet transform scalograms for the gearbox baseline vibration response obtained under; variable operating speed and constant load (VFCL), constant operating speed and variable load (CFVL), and variable frequency and variable load (VFVL) for helical and spur gear types are given in Table 13. The generated images are used as input for testing the proposed convolutional neural network model classification performance under
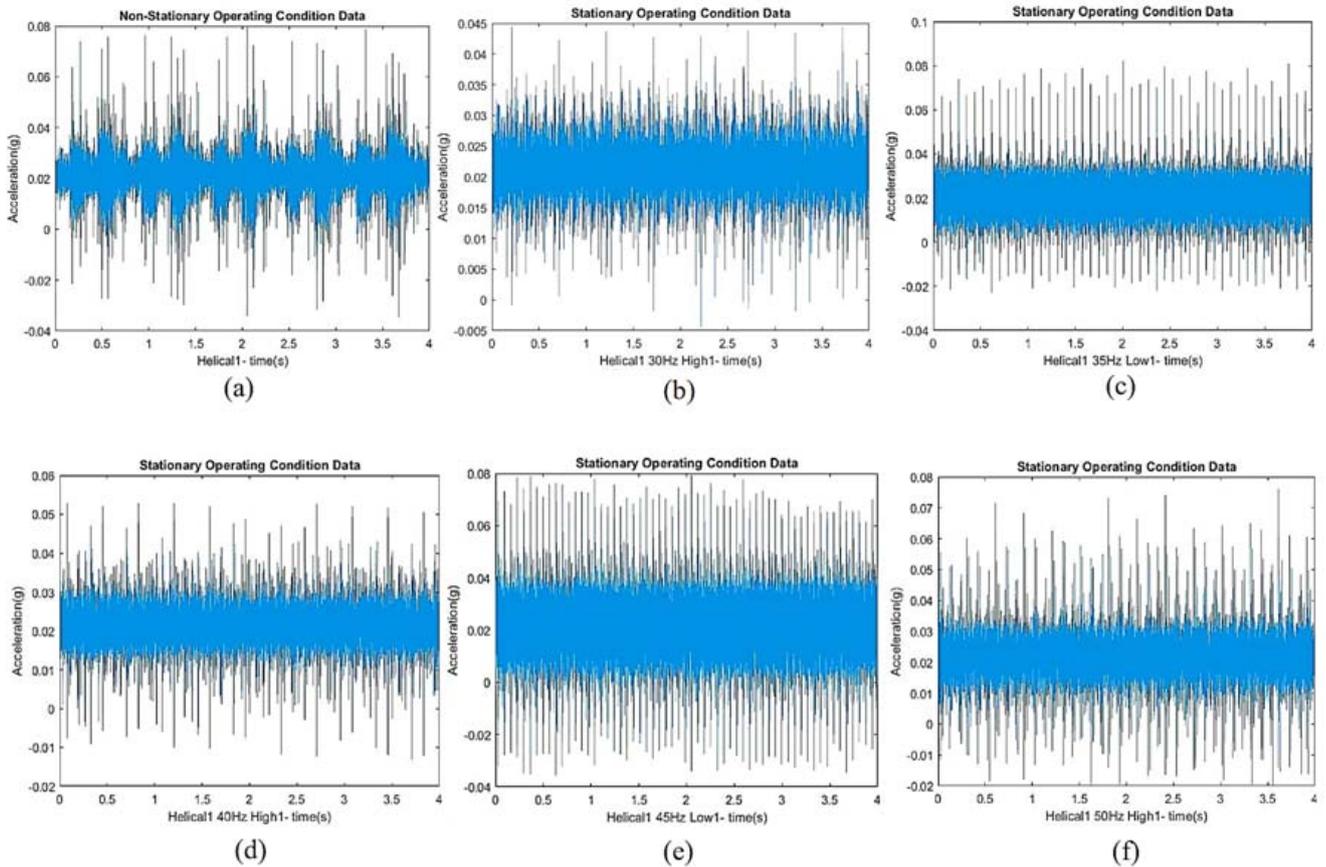
non-stationary operating conditions.

Figure 6. Time-domain representation of (a) Non-stationary operating conditions engineered vibration data sample, and stationary operating conditions vibration data samples obtained at an operating speed of (b) 30Hz under high loading, (c) 35Hz under low loading, (d) 40Hz under high loading, (e) 45Hz under low loading, (f) 50Hz under high loading.

Table 13. Non-stationary operating conditions continuous wavelet transform scalograms

Gear Type	Non-Stationary Operating Conditions Vibration Data Sets-Scalogram Images		
	Constant Operating speed and Variable Load (CFVL)	Variable Operating speed and Constant Load (VFCL)	Variable Operating speed and Variable Load (VFVL)
Helical			
Spur			

4.2. Developed Model Classification Evaluation under Non-Stationary Operations Conditions

Fault classification performance of the proposed convolutional neural network trained under stationary operating conditions is evaluated using engineered non-stationary operating conditions vibration data sets image representations. A gearbox operating under non-stationary conditions is subjected to variations in loading and operational speed, which in turn results in abrupt variations of vibration signatures as compared to a gearbox operating under stationary conditions under which the diagnosis model is developed. Therefore, the proposed model is evaluated to determine the extent to which a model developed under stationary operating conditions could be deployed as a fault diagnosis tool on machinery such as a gearbox operating under non-stationary conditions. The trained model performance evaluation results in multiple faults diagnosis of a two-stage gearbox using engineered non-stationary operating conditions vibration data sets is given in Table 14. The classification model is evaluated using time-frequency domain image representations of the vibration data sets obtained under the three non-stationary operating conditions (CFVL, VFCL, and VFVL) categories. The abbreviations given in Table 14, are defined as follows; developed model average classification accuracy (MACA) obtained using stationary operating conditions vibration data sets, model testing classification accuracy (MTCA) obtained using non-stationary operating conditions engineered vibration data sets.

Table 14. Developed model performance evaluation using non-stationary conditions image data input

Gear Type	Data Condition	Developed Model Classification Accuracy under Non-Stationary Operating Conditions									
		Time-Frequency Domain Representations Data Input									
		CWT		EMD		STFT		WSST		WVD	
		MACA	MTCA	MACA	MTCA	MACA	MTCA	MACA	MTCA	MACA	MTCA
Helical	CFVL		76.95%		68.03%		44.77%		63.88%		34.83%
	VFCL	99.67%	86.59%	98.11%	55.27%	96.94%	31.84%	98.37%	60.09%	98.05%	37.89%
	VFVL		73.63%		52.60%		33.07%		57.50%		35.22%
Spur	CFVL		84.91%		59.42%		47.24%		82.08%		46.42%
	VFCL	99.90%	83.69%	99.61%	56.93%	99.66%	27.54%	99.85%	82.08%	99.56%	49.90%
	VFVL		72.31%		47.51%		24.85%		83.45%		46.00%

The model highest classification performance under non-stationary operating conditions is obtained using scalogram (CWT, WSST) images data input, whilst the lowest performance is obtained using Wigner Ville plots and spectrograms data input. The classification model performance as high as 86% under non-stationary operating conditions using scalogram images gave the possibility that models developed under stationary operating conditions can be deployed successfully to diagnose faults in machinery operating under non-stationary conditions.

5. Model Validation under Stationary Operating Conditions

The model developed based on stationary operating conditions vibration data obtained from PHM2009 gearbox data set is validated using twenty percent validation data set obtained under stationary operating conditions, the developed model achieved good classification accuracy per each health pattern under spur and helical gear types and was validated using the five time-frequency (CWT, EMD, STFT, WSST, and WVD) image representations. The proposed model validation performance is given in Table 15. The scalograms achieved highest validation accuracy as compared other signal processing methods. The scalogram images (CWT and WSST) gave better performance due to the variation of scales of the Morlet wavelet as compared to the short-time Fourier transform with a fixed frequency resolution, which is dependent on the analysis window that consists of a fixed width.

Table 15. Model validation performance accuracies

Gear Type	Input Data Type				
	CWT	EMD	STFT	WSST	WVD
Helical	93.75%	84.38%	69.79%	91.13%	81.25%
Spur	94.53%	78.91%	77.34%	92.03%	91.14%

6. Conclusion

Gearbox fault diagnosis is a significant issue in the industry, therefore, early detection of faults in maintenance brings great economic value. Though the prior application of machine learning-based models in fault diagnosis were mostly limited by their sensitivity to experimental noise and their reliance on traditional feature extraction. The ability of the deep learning approach to automatically extract features from input data enables to speed up model development and mitigation of bias in feature engineering which is inherent in manual feature engineering. In this research work, the proposed convolutional neural network architecture was applied to five time-frequency image representations (CWT, EMD, STFT, WSST, and WVD) of raw vibration signals for gearbox multiple faults diagnosis. Faults diagnosis using the proposed architecture was achieved without the need for manual feature extraction and selection but was accomplished automatically signifying deep learning convolutional neural networks superiority in feature extraction.

The time-frequency signal representations were sized into four different image sizes (128×128, 96×96, 64×64, and 32×32) pixels, and the loss of information due to image scaling from 128×128 pixels to 32×32 pixels was found to have little effect on the scalogram images (CWT, WSST), as compared to the other image representations (spectrograms, EMD plots, and WVD plots). The proposed convolutional neural network architecture delivered nearly close accuracies for scalograms at lower image sizes of 32×32 pixels as compared to larger image sizes of 128×128 pixels, hence this slight difference in model performance and reduced training time could be exploited to reduce computational costs in diagnostic model development. The proposed convolutional neural network architecture also outperforms the other architectures under all the five time-frequency domain image representations (scalograms, spectrograms, Hilbert spectrum images, and Wigner Ville plots) input data. The proposed deep learning model achieved much higher performance accuracies as high as 99.90% as compared to the architectures which have been previously used in fault diagnosis. The use of image representation achieved higher classification as high as 99.90% as compared to manually engineered statistical feature vector data, which achieved highest accuracy of 82.91%. The results demonstrate that the proposed model can be accurately applied for multiple faults diagnosis based on time-frequency domain input data. Application of the trained model in multiple fault diagnosis under non-stationary operating conditions gave satisfactory results, as accuracies as high as 86.59% were obtained using CWT scalograms. The obtained results gave the possibility that the proposed model developed under stationary operating conditions can be deployed successfully to diagnose faults in machinery operating under non-stationary conditions.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] M. Muraro, F. Koda, U. Reisdorfer Jr, and C. H. D. Silva. (2012). "The influence of contact stress distribution and specific film thickness on the wear of spur gears during pitting tests." *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 34, pp. 134-144, 2012.

[2] O. Asi. (2006). "Fatigue failure of a helical gear in a gearbox." *Engineering Failure Analysis*, vol. 13, no. 7, pp. 1116-1125, 2006.

[3] A. S. Sait and Y. I. Sharaf-Eldeen. (2011). "A Review of Gearbox Condition Monitoring Based on vibration Analysis." *Rotating Machinery, Structural Health Monitoring, Shock and Vibration; Springer*, vol. 5, pp. 307-324, 2011.

[4] A. Aherwar and M. Khalid. (2012). "Vibration analysis techniques for gearbox diagnostic: a review." *Internal Journal of Advanced Engineering Technology*, vol. 3, no. 2, pp. 1-9, 2012.

[5] D. Verstraete, A. Ferrada, E. Droguett, V. Meruane, and M. Modarres. (2017). "Deep enhanced fault diagnosis using time-frequency image analysis of rolling element bearings." *Hindawi Shock and Vibration*, vol. 2017, no. 5067651, pp. 1-17, 2017.

[6] Z. Hou, Y. Zhang, P. Francq, S. J., and L. Huang. (2017). "Incipient fault diagnosis of roller bearing using optimised wavelet transform based multi-speed vibration signatures." *IEEE Acess*, vol. 5, pp. 19442-19456, 2017.

[7] J. Wang, J. Zhuang, L. Duan, and W. Cheng. (2016). "A multi-scale convolutional neural network for featureless fault diagnosis." In *Proceedings of the Internal Symposium on Flexible Automation*, Cleveland, Ohio, USA, August 2016.

[8]   M. Seera and C. Lima. (2014). "Online motor fault detection and diagnosis using a hybrid FMM-CART model." *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 4, pp. 806-812, 2014.

[9]   A. Sharma, A. Amarnath, and P. Kankar. (2014). "Feature extraction and fault severity classification in ball bearings." *Journal of Vibration and Control*, vol. 22, no. 1, pp. 1-17, 2014.

[10] P. Wong, J. Zhong, Z. Yang, and C. Vong. (2016). "Sparse Bayesian extreme learning committe machine for engine simultaneous fault diagnosis." *Neurocomputing*, vol. 174, pp. 331-343, 2016.

[11] Z. Chen, C. Li, and R. Sanchez. (2015). "Gearbox fault identification and classification with convolutional neural networks." *Shock and Vibration*, vol. 2015, no. 390134, pp. 1-10, 2015.

[12] Y. LeCun, Y. Bengio, and G. Hinton. (2015). "Deep learning." *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[13] C. Szegedy, W. Liu, and Y. Jia. (2015). "Going deeper with convolutions." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, Mass, USA, 2015.

[14] H. Liu, L. Li and J. Ma. (2016). "Rolling bearing fault diagnosis based on STFT-deep learning and sound signals." *Shock and Vibration*, vol. 2016, no. 6127479, pp. 1-12, 2016.

[15] L. Wang, X. Zhao, J. Wu, Y. Xie, and Y. Zhang. (2017). "Motor fault diagnosis based on shirt time Fourier transform and convolutional neural network." *Chines Journal of Mechanical Engineering*, vol. 30, no. 6, pp. 1357-1368, 2017.

[16] X. Guo, L. Chen, and C. Shen. (2016). "Hierachical adaptive deep convolutional neural network and its application to bearing fault diagnosis." *Measurement: Journal of the International Measurement Confederation*, vol. 93, pp. 490-502, 2016.

[17] K. Loparo. (2013). "Bearing Data Center,Case Western Reserve University, http://csegroups.case.edu /bearingdatacenter/pages/welcome-case-western-reserve-universitybearing-datacenter-website," 2013.

[18] F. Zhou, Y. Gao, and C. Wen. (2017). "A novel multimode fault classification method based on deep learning." *Jornal of Control Science and Engineering*, vol. 2017, no. 3583610, pp. 1-14, 2017.

[19] "Information on https://www.phmsociety.org/competition/PHM/09".

[20] S. Ioffe and C. Szegedy. (2015). "Batch normalisation:Accelerating deep network training by reducing internal covariate shift." *ArXiv*, vol. abs/1502.03167, pp. 1-11, 2015.

[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. (2015). "Drop out a simple way to prevent neural networks from overfitting." *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929-1958, 2015.