



# Bearing fault diagnostic using machine learning algorithms

Laith S. Sawaqed<sup>1</sup> · Ayman M. Alrayes<sup>1</sup>

Received: 24 February 2020 / Accepted: 19 September 2020 / Published online: 1 October 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

This study aims to enhance the condition monitoring of external ball bearings using the raw data provided by Paderborn University which provided sufficient data for motor current signal MCS. Three classes of bearings have been used: healthy bearings, bearings with an inner race defect, and bearings with outer race defect. Online data at different operating conditions, bearings, and faults extent of artificial and real damages have been chosen to provide the generalization and robustness of the model. After proper preprocessing to the raw data of vibration and MCS, time, frequency, and time–frequency domain features have been extracted. Then, optimal features have been selected using genetic algorithm. Artificial neural network with optimized structure using genetic algorithm has been implemented. A comparison between the performance of vibration and motor current signal has been presented. Moreover, our results are compared to previous work by using the same raw data. Results showed the potential of motor current signal in bearing fault diagnosis with high classification accuracy. Moreover, the results showed the possibility to provide a promised diagnostic model that can diagnose bearings of real faults with different fault severities using MCS.

**Keywords** Bearing damage detection · Machine fault diagnostic · Vibration · Motor current signal · Machine learning algorithm · Neural networks · Genetic algorithm

## 1 Introduction

There is no doubt that the proper maintenance management system contributes significantly in increasing the company profits and prevents huge losses. Condition monitoring (CM) and predictive maintenance are an essential aspects for improving a sufficient maintenance management system, for instance; a case study for a paper mill at Swedish showed that preventing unplanned stoppage for a year will increase the profits by 0.975 million USD [1]. CM is essential to avoid harmful consequences and reduce financial loss and that's why it has been attracting the interest of researchers for the last few decades [2].

Machine failure occurs due to many reasons, such as stator faults, rotor faults, or bearing faults [3]. Statistics, in the industrial world, showed that up to two-thirds of motor fail-

ures in the electromechanical drive systems are initiated in the bearings due to rolling bearing damages [4].

Data-driven bearing fault diagnostic passes through three main steps to give a final decision regarding the state of the bearing as per ISO 13374, which are data acquisition (DA), data manipulation (DM), and state detection (SD). In the first step, raw data are to be collected from the system. These data could be any signal that carries the signature of the fault, for instance, vibration signal, acoustic emission, or motor current signal (MCS). In the second step, the acquired signal is processed to gain helpful information about the bearing under monitoring. This step could be named data analysis or signal processing step. It starts with signal preprocessing, feature extraction and selection. A variety of data processing techniques have been used by researchers, namely time domain analysis [5–7], frequency domain analysis [8, 9], and time–frequency domain analysis [10]. Soualhi et al. [11] presented an excellent survey for signal processing techniques that could be used in condition monitoring for bearings and gearboxes. The third step is the decision-making, where the selected features of the previous step may be analyzed to decide whether the bearing is faulty or not, or could be intro-

---

✉ Laith S. Sawaqed  
lssawaqed@just.edu.jo

Ayman M. Alrayes  
eng.ayman.alrayes@gmail.com

<sup>1</sup> Mechanical Engineering Department, Jordan University of Science and Technology University, Irbid, Jordan

duced to diagnostic model which in turn classifies the bearing as healthy or faulty bearing [12].

In the last two decades, the emergence of the science of artificial intelligence (AI) in bearing condition monitoring and fault diagnostic provided automated algorithms capable to give final diagnostic decision or even to diagnose and classify the different types of bearing faults. This has contributed effectively in minimizing time, effort, and cost in the fault diagnostic process. Many AI techniques have been used by researchers in bearing fault detection such as pattern recognition, fuzzy logic [13], expert system [13], neural network (NN) [6, 7, 14, 15], neuro-fuzzy system [8], genetic algorithm (GA) [6, 15], support vector machine (SVM) [7, 16], and many others.

Paper [14], to the knowledge of the author, was the first who employed artificial neural network (ANN) to diagnose rolling element bearing defects using the frequency spectrum of vibration signal with recirculation neural network. Their technique enhanced traditional vibration analysis and provided an automated system to monitor and diagnose bearing rolling element defects.

On the other hand, Paper [15] used the genetic algorithm (GA) in order to choose the optimal input features to ANN, where choosing the best features is considered as a challenge in neural network; too many features result in long computational time and sometimes with lower performance results than using less features with optimal performance. In their study, the performance of six selected features was better than the performance of full features set by 13%.

Paper [6] studied the effect of the number of input features, and the signal preprocessing prior to the feature extraction process on the training performance of the ANN and its effectiveness in diagnosing bearing defects. Time domain features were extracted from segments of vibration signal which in turn was acquired from two bearings: one healthy and the other of outer race defect. Features were then presented to ANN of two hidden layers and two outputs of two bearing health classes. Their study showed the importance of the type of features on the ANN performance and also the substantial effect of a small number of features on ANN computation time and testing time. The main drawback of using ANN in machine fault diagnostics is the difficulty to be generalized for other machines and should be retrained using new signals once the machine is changed.

Paper [7] designed a classifier using two machine learning algorithms ANN and SVM where time domain features were extracted from vibration signal. Five types of bearings were classified: healthy bearing, outer race crack, inner race rough surface, ball with corrosion pitting, and combined defects. Generally, the study showed the potential application of machine learning algorithms in early diagnosis of bearing defect.

Paper [8] used the frequency domain of vibration signal in bi-classifier approach to classify bearing faults, where several neuro-fuzzy systems were used in cascade. The first system classifies the bearings into healthy or faulty bearing. The second system classifies the faulty bearings into a ball or race defect. Results showed that neuro-fuzzy systems with only two outputs have better performance. However, it was difficult to distinguish between inner and outer race defects due to the similarity in the vibration spectral.

Mostly, researchers relied on the vibration signals in bearing fault diagnostic, but not that many used the MCS. As the vibration signal proved its ability and reliability to detect the bearing damage over decades, the MCS still has restrictions [17]. For instance, the accelerometer or vibration detector is placed directly on the bearing under study to reduce nearby noise that may affect the vibration signal, while the transfer of the defect signature over drive train and motor components may affect the MCS by increasing the chance to get lost or masked by a noise and suffering from disturbances [18]. Also, the diversity and intensity of the available vibration data make the studies easier, while the lack of availability of MCS limits further research [19, 20]. Nevertheless, vibration analysis has a drawback since the location of the bearing in the machine is not always easily accessible. On the other hand, vibration analysis is relatively expensive compared to MCS since no sensors installation and wiring are needed. MCS can be acquired directly from the frequency inverters. Paper [17] provided a systematic data set of disciplinary bearings of real faults and artificially induced faults. The data were recorded synchronously for vibration and MCS under different operation conditions.

This work contributes by proving the feasibility of using the data acquired for MCS to be used for bearing fault diagnostic with high performance accuracy. In addition, the use of bearings with different types of faults provided robust and reliable bearing fault diagnostic model by using ANN and genetic algorithm based on the vibration and MCS data provided by [17].

## 2 Bearing defects

Many criteria have been used to classify bearing faults by researchers based on research needs. One popular criterion is the bearing damage location, where the bearing faults are classified into inner ring fault, outer ring fault, rolling element fault, and race fault [21]. Another criterion based on the damage combination classifies the bearing damages into single point defect which include the four damages mentioned previously, and generalized roughness or distributed damages, where the damage of the bearing is not apparent to the unaided eye, or can't be characterized by a unique frequency [22]. In this work, a classification model has been presented

to diagnose bearings whether it's healthy or of single point defect, mainly inner race defect and outer race defect.

In a single point defect faulty bearing, once the bearing runs over the crack or flaw of the fault, the bearing responds by an impulse or “ringing” frequency decays in a short period, this frequency usually above 5 kHz [23]. Moreover, each element in the bearing, inner ring, outer ring, ball, and train has its own characteristic frequency, Eqs. (1)–(4), which depends on the bearing geometrical parameters, and once a fault is present in one element, the induced impulse increases the energy of the acceleration spectrum at the corresponding faulty element characteristic frequency [24].

$$f_{IR} = \frac{n}{2}N \left( 1 + \frac{BD}{PD} \cos \varnothing \right) \quad (1)$$

$$f_{OR} = \frac{n}{2}N \left( 1 - \frac{BD}{PD} \cos \varnothing \right) \quad (2)$$

$$f_{BD} = \frac{PD}{2BD}N \left( 1 - \left( \frac{BD}{PD} \right)^2 \cos^2 \varnothing \right) \quad (3)$$

$$f_{TD} = \frac{1}{2}N \left( 1 - \frac{BD}{PD} \cos \varnothing \right), \quad (4)$$

where  $n$  is the number of balls,  $N$  is the shaft rotation frequency (Hz),  $bd$  is the ball diameter (mm),  $PD$  is the pitch diameter (mm), and  $\varnothing$  is the contact angle in radians.

On the other hand, any fault in the bearing will affect the motor supply current. In [25], a study for the relationship between the vibration and current frequencies due to bearing defects was presented. It stated that the presence of bearing defect produces a slight change in shaft rotation. This influences the air gap producing air gap eccentricity which will cause a variation in the air gap flux density, which in turn will influence the stator current by generating a stator current at frequencies given by:

$$f_{bng} = |f_e \pm m \cdot f_v|, \quad (5)$$

where  $f_{bng}$  is the generated predictable frequency (Hz),  $f_e$  is the electrical supply frequency (Hz),  $m$  is an integer 1, 2, 3 ..., and  $f_v$  is the vibration characteristic frequency of the corresponding bearing defect (Hz).

### 3 Data collection

In this work, the data provided by Paderborn University, which are available online [26], were chosen for this study due to its variety of bearing damages, validity, comprehensive data-driven using machine learning, and having both vibration and MCS. As mentioned earlier one of the goals

of this work is to make the MCS applicable for condition monitoring of bearing damage due to its financial feasibility. To generate the systematic data for different healthy bearings or faulty bearings of different severities, a test rig of Fig. 1 was designed at the Chair of Design and Drive Technology, Paderborn University. The test rig was operated to generate the measurement data for electrical motor current and bearing vibration using current transducer and accelerometer respectively. The data were acquired for 4 s and sampled using ADC at 64 kHz. For more details about the experiment, please refer to [17].

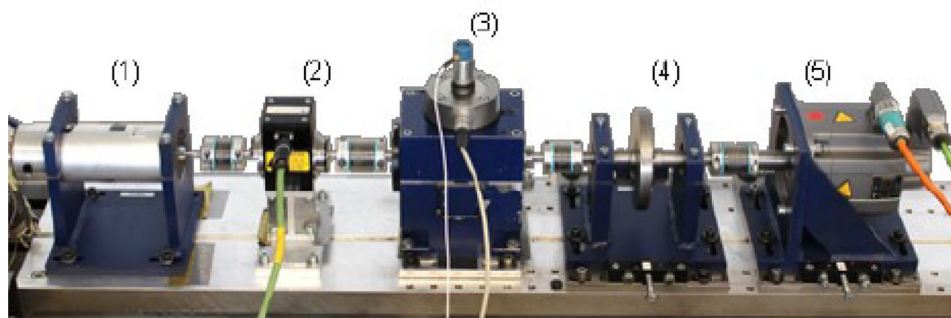
The test rig was operated for several times, each time with different bearings; all bearings used in this experiment to derive systematic data were ball bearings with eight rolling elements of type 6203 with nearly same geometrical sizes. Healthy and faulty bearings were used. Faulty bearings are of inner ring fault and outer ring fault, real and artificially induced damages which added more reliable and robustness to the diagnostic model. To differentiate between different bearings in measurement records, healthy bearings are denoted by a code starting with “KO,” inner ring faulty bearings start by the letters “KI,” while the outer ring bearings start with “KA.” A summary of the bearings selected in this study with the fault method generation is shown in Table 1.

Artificially induced damages in bearing fault studies are introduced manually in the laboratories by different methods. Inner race, outer race, and even cage defects could be produced.

Some researchers based their research on outer race only [27], as the outer ring faults are easier to be detected than inner ring faults because the outer ring is the stationary part of the bearing while the inner ring is the rotating part [28]. Other researchers did their work based on artificial damages only without involving real damages [13, 18]. What makes the data of [17] distinct is its inclusion on outer, inner, real, and artificial damages even with different severities, extent, and combination of damages.

Vibration and phase current measurements were recorded on different operating conditions. Four operating conditions were used to provide various data denoted by S0, S1, S2, and S3. Each operating condition has three operating parameters, namely rotational speed, load torque, and radial force. Each time one of three operating parameters was changed. Table 2 shows the different operating conditions used to record the data with setting name, for each time one operator condition is changed. The last column of the table shows setting names which were used to save the data as MATLAB files. Twenty measurements each of 4 s length, sampled at 64 kHz, were recorded for each operation condition. And as mentioned earlier, a total of 29 bearings have been used in this work.

**Fig. 1** Modules of Paderborn test rig for bearing data acquisition (1) electric motor, (2) torque measurement shaft, (3) rolling bearing module, (4) flywheel, and (5) load motor [17]



**Table 1** Summary for the number of bearings used in this study

Category	Generating	Amount
Inner ring fault (IR)	Real	6
	Artificial	5
Outer ring fault (OR)	Real	5
	Artificial	7
Healthy bearing		6
Total		29

### 4 Signal preprocessing

The idea of preprocessing is to facilitate the neural network training using different steps applied to the raw data, like signal filtration, normalization, and feature extraction. These steps are important in input data preparation to the neural network (NN), so it becomes easier to NN to extract relevant information [29].

In this work, the data are sampled at 64 kSa/s for 4 s, while the shaft is operated at speeds of 1500 or 900 rpm which means 25, and 15 Hz, respectively. In comparison with the literature, the data used here are of long duration and high frequency. However, many researchers in the literature divided the raw signal into many segments either to increase the number of extracted features or to divide the same signal into training and testing data due to lack of raw signals [8, 27]. In this work, due to the diversity and availability of different bearings of different faults, severities, and generation method including real and artificial, also each bearing has a 20 measurement, no need to divide the same raw signal into training and testing data. Bearings of artificial damages were used for training, whereas those of real damages were used

for testing. The raw signal was denoised first and then segmented, and the features were extracted from one segment for each measurement; the features were normalized next, and suitable features have been chosen by using GA.

#### 4.1 Signal segmentation

The signal has been segmented after denoising the raw data using discrete wavelet transform (DWT). Our data have in each single measurement a total of 256 k sample recorded during the 4 s, while each shaft rotation cycle duration is 40 ms when rotating at 1500 rpm and about 70 ms when rotating at 900 rpm. This means that the data were recorded for 100 and 60 shaft cycles for the two shaft speeds, respectively. Raw data have been segmented into five shaft cycles. In other words, 12,800 samples have been considered for measurements of 1500 rpm and 21,333 samples have been considered for measurements of 900 rpm shaft speed. One segment of each measurement has been considered for feature extraction. Segmentation is important to reduce the computation time by eliminating redundant data.

#### 4.2 Feature extraction

The examination of wide range of features was proved to be useful in bearing fault diagnostic, where the feature value depends on the bearing fault type and its severity. For instance, time domain features are very useful when bearing fault signal is highly impulsive. However, it becomes ineffective when the fault is of high severity or when the bearing is overloaded. On the contrary, frequency domain parameters are more effective to detect bearing faults due to overload or faults with high severities. Based on a review, it was stated

**Table 2** Different operating conditions with setting names

No	Rotational speed (rpm)	Load torque (Nm)	Radial force (N)	Setting name
$S_0$	1500	.7	1000	N15_M07_F10
$S_1$	900	.7	1000	N09_M07_F10
$S_2$	1500	.1	1000	N15_M01_F10
$S_3$	1500	.7	400	N15_M07_F04



that no exclusive parameters could be used to detect bearing faults; instead, multi-parameter tests were recommended to be used in bearing fault detection [30].

Moreover, because the characteristic frequencies of localized defects are of low energy, it could be easily masked by the vibration of another structural component. This makes detecting bearing defects using frequency spectrum sometimes challenging. Therefore, non-stationary signal analysis, wavelet transform (WT), provided a great solution to overcome the non-stationary nature of faulty bearings signals [31]. Nonstationary term refers to frequency changing with time.

Because in this work, the data were collected for multi-disciplinary bearings, faults, severities, and operation conditions, the features were extracted in different domains including time, frequency, and time–frequency domains. Eleven time domain features were extracted using statistical operations on processed signals, which are peak value, RMS, variance, max value, crest factor, skewness, kurtosis, Shannon entropy, clearance factor, impulse factor, shape factor, and peak-to-peak value. Two frequency domain features were extracted from frequency spectrum by performing fast Fourier transform (FFT) analysis on time domain signal, peak value of FFT, and power spectral density of FFT. Wavelet packet decomposition (WPD) has been employed to extract the time–frequency domain features. WPD decomposes the signal into multiple levels. Each level has twice as many signals as the previous one, which means each signal is decomposed into two signals. Starting from the first signal, two signals are generated from it: one comes from filtering the signal using low-pass filter and called approximation signal (A). The second comes from filtering the signal using high-pass filter and called detailed signal (D) [32]; in our case, the time domain signal was decomposed into three levels using bior3.7 wavelets. Then, the energy of approximate coefficient 3 ( $cA3$ ) and detailed coefficients ( $cD$ ) 1, 2 and 3 have been obtained. A total of 17 features have been extracted. More explanation and formulas of the features are presented in [33].

### 4.3 Normalization

In multilayer NN, a sigmoid activation function is used for the hidden layers. Sigmoid functions become saturated when the inputs become larger than 3 by nature as it is an exponential function [29]. Furthermore, some features span a range much wider than other features, which means some feature vectors have tiny numbers while other vectors may have larger values of two orders of magnitude. For these reasons, a normalization process is performed for input data vectors. All features have been normalized to span the space 0 to 1. Only skew-

ness was normalized to span the space - 1 to 1. Equation (6) was used for normalization.

$$P^n = \frac{P - P^{\min}}{P^{\max} - P^{\min}} \quad (6)$$

where  $P^n$  is the resulting normalized input vector.  $P$  represents the non-normalized input vector.  $P^{\min}$  is a vector of minimum values of each element of the input vector, while  $P^{\max}$  is a vector represents the maximum values of each element of the input vector of input data set.

## 5 Methodology

Once the best features are selected by forward selection process (i.e., testing each feature individually and then making combinations from those features to have best performance) to train the ANN, if two features acting individually poor, but when used together they may give a much better result than the two best features achieved through forward selection [15]. Besides, many authors discussed and proved the importance of feature selection in bearing fault diagnostic [6, 15]; the full feature set was tested first in this work showing very low performance accuracy comparing to the results with feature selection. Testing with complete feature set didn't exceed 60% performance accuracy. This makes the GA an excellent choice for feature selection process. Many authors discussed the importance of using GA for feature selection process as shown in [27, 34]. Moreover, the tuning of ANN hidden layers is also a coherent challenge in ANN, where it depends only on trial and error. Trying manually hundreds of trials to find the best network structure is not timely feasible. GA was used to overcome this challenge, where generations of individuals could be tested; each individual consists of combination of the number of neurons in the two hidden layers. Depending on the number of individuals within the generation and maximum number of generations, hundreds and even thousands of NNs could be trained and saved each time with different NN structure. Then, the best performance NN could be chosen for the classification model [27].

In this work, hybrid system of ANN and GA is implemented to simulate the classification model, where the GA was used twice, first time for feature selection and second time to tune the ANN structure. Features were used as inputs to the classification model, while a target matrix representing three different classes was prepared to be used as an output. Data of artificial damages were used for training, and data of real damages were used to test the trained ANN.

## 5.1 Artificial neural network

Two-hidden-layer NN was used to perform the pattern recognition for our classification problem. The number of hidden layers was selected based on trial and error, where the neural network was trained on 2, 3, and 4 hidden layers. For 3 and 4 layers the training process took a very long computational time without performance improvement, taking into account the computational time of genetic algorithm during this training trials to tune the number of neurons in the hidden layers also increasing vastly. This was compatible with what is presented in the literature where two hidden layers were chosen [6].

The number of neurons in each layer was determined by the GA as described in the next section.

Our problem is to diagnose the bearing through its signal and classify it either as healthy bearing, inner ring-defected bearing, or outer ring-defected bearing. Inputs used are the selected features and extracted from the processed vibration signal or MCS. The number of inputs is equal to the number of selected features. Three outputs were used to represent the three classes. For each class one output is set to 1 and the other two are 0 s. In such pattern recognition problems, a SoftMax output layer is used, such that one output will be set to one and all others are zero. This could be achieved using a probability distribution function as an activation function for the output layer [29].

## 5.2 Genetic algorithm

In this work, GA has been applied twice, first time to select the optimal features from a complete extracted feature set and the second time to tune the neuron number of the hidden layers of the ANN.

## 5.3 Chromosome representation

For feature selection, the chromosome ( $X$ ) is encoded by the number of the selected features ( $N$ ) from the feature set of ( $R$ ) features. So, the chromosome string consists of  $N$  real numbers ( $X_i; i = 1 : N$ ) and each real number of  $X_i$  is bounded by the range ( $1 \leq X_i \leq R$ ).

For ANN training, the chromosome ( $X$ ) is encoded by the number of neurons of each hidden layer, supposing ( $N$ ) hidden layers. So, the chromosome string consists of  $N$  real numbers ( $X_i; i = 1 : N$ ) and each real number of  $X_i$  is bounded by  $S_{\min}$  and  $S_{\max}$ , where  $S_{\min}$  and  $S_{\max}$  are the lower and upper bounds of neurons for each of the hidden layer.

$$S_{\min} \leq X_i \leq S_{\max} \quad (7)$$

$S_{\min}$  and  $S_{\max}$  are 2 and 20, respectively. The resulted NN structure, i.e., number of neurons in each layer, was different for each operating condition.

## 5.4 Initial population, fitness function, and termination

The population is the number of individuals in each generation which consists of *population size* of rows and  $N$  columns of problem variables [35]. An initial population has to be defined to the GA in order to start the optimization process. Random generation method is used in this work as it is an integer problem [35], where a population size of 20 individuals was generated based on the chromosome selection limits discussed earlier.

Next generations were produced continuously by selection, mutation, and crossover operations to the parents. Each individual is assessed by the fitness function. Generations will be produced continuously until a stopping criterion is satisfied. The most common termination criterion is the maximum number of generations, which is used in this work such that the GA is set to produce a maximum number of 51 generations.

The fitness function, aka objective function, is used to assess the performance of an individual compared to other individuals, called the fitness score. The individual fitness score determines which individuals will be selected to produce the next generation. Herein, the performance of ANN to classify the testing data correctly was used as a fitness function. The less misclassification samples result in higher fitness index.

## 5.5 Genetic operators

Two genetic operators have been applied in this work, namely mutation and crossover. Crossover operation combines two parents to produce a child of next generation. Many crossover operators could be used based on the type of chromosome and application of GA. Many types of crossover operators were reviewed to be helpful for researchers [36]. In this work, as our problem is of integer constraints problem, then Laplace crossover has been implemented [37]. It is worth mentioning that crossover operation should ensure that new generations satisfying the bounds and constraints.

On the other hand, mutation operation makes random changes in the individuals of the population to create children. This provides genetic diversity to search within the solution space [35]. Furthermore, mutation enables the optimization process to avoid local minima values. Because our problem is an integer problem, i.e., both features vector and number of neurons are integers only, then selecting the mutation function should be done carefully, as the mutation may result in a population not necessarily satisfying the

**Table 3** List of data used for testing and training

Class	Class name	Training	Testing
1	Healthy	K001	K004
		K002	K005
		K003	K006
2	IR	KI01	KI04
		KI03	KI14
		KI05	KI16
		KI07	KI17
		KI08	KI18
3	OR	KA01	KA04
		KA03	KA15
		KA05	KA16
		KA06	KA22
		KA08	KA30

constraints and the bounds. Herein, the power mutation is adopted [37, 38]. For generations reproduction, 5% of the generation are selected as an elite children, which will be survived as they are the fittest, 80% of the rest of the new generation are generated by crossover operation from the previous generation, and the remaining individuals are mutation children.

## 6 Results and discussion

The model was run for vibration data first four times, each time for one operation condition. Each run on each operation condition is completely independent of the others. This means optimal features and NN structure are different for each operation setting. Then same procedure is repeated for MCS data. After that, the model was run for selected data to compare the results of the model with other studies using the same raw data.

### 6.1 Training and testing using vibration data

The model was tested for the four operation conditions;  $S_0$ ,  $S_1$ ,  $S_2$ , and  $S_3$  using vibration data in order to investigate the effect of different operating conditions on classification model and feature selection. The data used for training and testing are listed in Table 3.

Based on Table 3, three healthy bearings were selected for testing, which means 60 examples, as each bearing includes 20 measurements as discussed in the data section 0. Also, 6 bearings were chosen as IR, which means 120 samples, 100 samples presented the OR as well. Same data partitioning of

**Table 4** Confusion matrices for classification model trained with artificial damages and tested with real damages using vibration data for different operating conditions

Target	H			IR			OR			All (%)
Output	H	IR	OR	H	IR	OR	H	IR	OR	
$S_0$	60	0	0	0	113	7	22	14	64	84.6
$S_1$	59	1	0	4	113	3	1	2	97	96.1
$S_2$	51	0	0	7	113	0	2	0	98	93.6
$S_3$	56	4	0	17	101	2	0	3	97	90.7

**Table 5** Confusion matrices for classification model trained with artificial damages and tested with real damages using MCS data for different operating conditions

Target	H			IR			OR			All (%)
Output	H	IR	OR	H	IR	OR	H	IR	OR	
$S_0$	43	1	16	0	120	0	4	0	96	92.5
$S_1$	54	2	4	2	95	23	0	1	99	88.6
$S_2$	41	0	19	0	120	0	8	1	91	90.0
$S_3$	94	3	8	6	110	4	2	1	97	91.4

Table 3 was used for the four operating conditions  $S_0$ ,  $S_1$ ,  $S_2$ , and  $S_3$ .

By using the SoftMax output layer in neural network, all inputs will have a classification. This implies that the misclassification happening in certain data will be classified to one of the other two classes. Table 4 shows the correct classified samples and misclassification occurred while testing the classification model for the selected data.

It is clearly shown from Table 4 that the best overall classification result using vibration signal is for operating condition  $S_1$ , where it reached 96.1%. Also it is the best to classify the faulty as faulty (regardless IR or OR) where 5 faulty samples only out of 220 were classified as healthy. While the best true positive result is for operating condition  $S_0$ , that is all healthy samples were classified correctly as healthy.

### 6.2 Training and testing using MCS

Same procedure of vibration data was repeated using MCS, where the MCS features of same data set were trained and tested using the same classification model for all the four operating conditions  $S_0$ ,  $S_1$ ,  $S_2$ , and  $S_3$ . Table 5 shows the confusion matrices for each operating condition. From the table, it is clearly shown that the best performance of MCS was using the data of operating condition  $S_0$  with a classification accuracy of 92.5%.

**Table 6** List of data sets used for comparison with work of paper [17]

Class	Class name	Training	Testing
1	Healthy	K002	K002
2	IR		KI14
		KA01	KI21
		KA05	KI17
		KA07	KI18 KI16
3	OR		KA22
		KI01	KA04
		KI05	KA15
		KI07	KA30
			KA16

### 6.3 Comparison between vibration and MCS

Although the MCS showed good results for bearing fault detection, but still the vibration signal showed better results. Moreover, one important parameter has to be considered in such classification models, that is the misclassification occurred between healthy and faulty bearings in general. False positive are healthy samples classified as faulty, while false negatives are faulty samples classified as healthy. Comparing vibration and MCS results in overall samples classification performance for different operating conditions; vibration signal always shows better false positives even when the overall performance for the operating condition is less accurate; i.e., it is better to classify the healthy bearings as healthy not faulty. This is important in reality because it will result in an incorrect maintenance decision to change the healthy bearings in the electromechanical systems, as it is diagnosed as faulty, which means a financial loss.

### 6.4 Comparison with other studies

The classification model has been implemented on the same data set used by [17] where a certain group of data set was used in their work to train various machine learning (ML) algorithms using features of artificial damages and tested using features of real damages. Seven ML algorithms have been used in their work. One of those algorithms was the NN, which is the main ML classifier used in this work. The comparison presented here is between the results of NN performance of their study and the NN performance of this study. Table 6 shows the bearings used by [17] which is also used in this work for comparison purpose. The data were extracted from operating condition 0 only;  $S_0$ .

Table 7 shows the comparison between the results of the previous work and the results of this work for the same data set. It is worth mentioning that the procedure of [17] is totally

**Table 7** Classification accuracy comparison for selected bearings of operating condition  $S_0$  (comparison between [17] results and this work for same data set)

Method	Classification accuracy	
	Vibration (%)	MCS (%)
NN [17]	65.5	45.5
Ensemble 7 ML algorithms [17]	75	45.9
This work proposed method	94.1	95

different in feature selection and ML algorithm training. Maximum separation distance method was employed for feature selection, and ML algorithms were trained without any tuning. However, in this work, GA optimization approach used for feature selection. Moreover, the ANN was tuned by GA as well. Table 7 also shows the ensemble of 7 classifiers used by the study of [17]. Ensemble learning combines several models of machine learning to provide a predictive model with performance better than single model. Moreover, [39] implemented the state of the art deep neural network (DNN) using adversarial auto-encoder as a main structure of the DNN. Hilbert–Huang transform (HHT) was adopted for data preprocess; then, the low-frequency components of both vibration and MCS were retained for analyzing. The current and vibration signals fused to detect whether there is inner ring fault, outer ring fault or no fault using the same partitioning of data set presented in [17] which is under comparison in this section. Their model distinguished the fault type with the accuracy of 85%.

## 7 Conclusion and future work

Machine learning algorithm of neural network has been applied efficiently to classify three categories of bearings of different health statuses: healthy bearings, inner ring fault, and outer ring fault. GA has been applied to find optimal features and optimize ANN structure. The results of this work show that the MCS has a great potential to diagnose the external bearing fault in electromechanical drive systems with high classification accuracy performance. This led to the possibility of providing a promising bearing fault diagnostic model that can monitor the external bearings without any additional costs for sensor installation which makes bearing fault diagnostic systems financially feasible for industry.

The model has been tested for different operating conditions using different bearings of real damages with different severities. Most classification results exceeded 90% performance for both vibration or MCS. This provides robustness model can be used in reality to diagnose the status of the bearing for one of the three classes treated in this work regardless the severity of extent of the damage. This is considered to



be superior to previous studies which were restricted to relatively limited data set used in the study itself.

Even though the classification model in this thesis showed relatively good results, the author believes that better results could be obtained by examining different CM techniques or applying more efforts on signal processing to extract more usable features, specially to decrease the percentages of false positives and false negatives using MCS.

Moreover, in this work, the machine learning algorithm model has been implemented for each operating condition separately. More generalized model of cascaded NN or deep learning algorithm could be presented to use the data of all operating condition at once, so the model can recognize the operating condition and bearing fault. Such model will be much more general and compatible to the industry where the faults are required to be recognized regardless the operating condition.

**Acknowledgements** The authors would like to acknowledge the Mechatronics and Dynamics research group at the University of Paderborn for offering an internship opportunity to work in their labs and projects through NRW scholarship program (<https://mb.uni-paderborn.de/en/ldm/>). Special thanks to Dr. James Kuria (jki-motho@campus.upb.de) for his valuable guidance, thoughts, and discussion which helped to present this work.

## References

1. Alsyouf, I.: The role of maintenance in improving companies' productivity and profitability. *Int. J. Prod. Econ.* **105**, 70–78 (2007)
2. Al-Najjar, B., Wang, W.: A conceptual model for fault detection and decision making for rolling element bearings in paper mills. *J. Qual. Maint. Eng.* **7**(3), 192–206 (2001)
3. Henao, H., et al.: Trends in fault diagnosis for electrical machines: a review of diagnostic techniques. *IEEE Ind. Electron. Mag.* **8**(2), 31–42 (2014)
4. Bonnett, A.H., Yung, C.: Increased efficiency versus increased reliability. *IEEE Ind. Appl. Mag.* **14**(1), 29–36 (2008)
5. Wu, D., Wang, H., Liu, H., He, T., Xie, T.: Health monitoring on the spacecraft bearings in high-speed rotating systems by using the clustering fusion of normal acoustic parameters. *Appl. Sci.* **9**(16), 3246 (2019)
6. Samanta, B., Al-Balushi, K.R.: Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mech. Syst. Signal Process.* **17**(2), 317–328 (2003)
7. Kankar, P.K., Sharma, S.C., Harsha, S.P.: Fault diagnosis of ball bearings using machine learning methods. *Expert Syst. Appl.* **38**(3), 1876–1886 (2011)
8. Marichal, G., Artes, M., Garcia-Prada, J.: An intelligent system for faulty-bearing detection based on vibration spectra. *J. Vib. Control* **17**(6), 931–942 (2011)
9. Khadersab, A., Shivakumar, S.: Vibration analysis techniques for rotating machinery and its vibration analysis effect techniques for rotating on bearing faults machinery and its effect on bearing faults costing models for capacity optimization in industry 4.0: trade-off between used capacity and operational efficiency. *Procedia Manuf.* **20**, 247–252 (2018)
10. Nikolaou, N.G., Antoniadis, I.A.: Rolling element bearing fault diagnosis using wavelet packets. *Ndt & E Int.* **35**, 197–205 (2002)
11. Soualhi, A., Hawwari, Y., Medjaher, K., Guy, C., Razik, H., Guillet, F.: PHM survey: implementation of signal processing methods for monitoring bearings and gearboxes. *Int. J. Prognostics Health Manage.* **9**, (2018)
12. Jardine, A.K.S., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **20**, 1483–1510 (2006)
13. Liu, T.I., Singonahalli, J.H., Iyer, N.R.: Detection of roller bearing defects using expert system and fuzzy logic. *Mech. Syst. Signal Process.* **10**, 595–614 (1996)
14. Alguindigue, I.E., Loskiewicz-Buczak, A., Uhrig, R.E.: Monitoring and diagnosis of rolling element bearings using artificial neural networks. *IEEE Trans. Ind. Electron.* **40**(2), 209–217 (1993)
15. Jack, L.B., Nandi, A.K.: Genetic algorithms for feature selection in machine condition monitoring with vibration signals. *IEE Proc. Vis. Image Signal Process.* **147**(3), 205–212 (2000)
16. Jenkins, C.D.: Bearing fault detection and wear estimation using machine learning. Los Alamos National Lab (LANL), Los Alamos (2019)
17. Lessmeier, C., Kimotho, J.K., Zimmer, D., Sextro, W.: Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification. In: European Conference of the Prognostics and Health Management Society 2016, no. July, p. 17 (2016)
18. Blodt, M., Granjon, P., Raison, B., Rostaing, G.: Models for bearing damage detection in induction motors using stator current monitoring. *IEEE Trans. Ind. Electron.* **55**(4), 1813–1822 (2008)
19. Zarei, J., Poshtan, J.: An advanced Park's vectors approach for bearing fault detection. *IEEE Int. Conf. Ind. Technol. Mumbai* pp. 1472–1479 (2006). <https://doi.org/10.1109/ICIT.2006.372562>
20. Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., Varnier, C.: PRONOSTIA: an experimental platform for bearings accelerated degradation tests. *IEEE Int. Conf. Prognostics Health Manag. PHM'12*, Denver, Colorado, United States. pp.1–8. (hal-00719503) (2012)
21. McFadden, P.D., Smith, J.D.: Model for the vibration produced by a single point defect in a rolling element bearing. *J. Sound Vib.* **96**(1), 69–82 (1984)
22. Stack, J.R., Habetler, T.G., Harley, R.G.: Fault classification and fault signature production for rolling element bearings in electric machines. In: *IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, SDEMPED 2003—Proceedings*, vol. 40, no. 3, pp. 172–176 (2003)
23. McNerny, S.A., Dai, Y.: Basic vibration signal processing for bearing fault detection. *IEEE Trans. Educ.* **46**(1), 149–156 (2003)
24. Tondon, N., Choudhury, A.: A review of vibration and acoustics measurement methods for the detection of defects in rolling element bearing. *Tribol. Int.* **32**(8), 469–480 (1999)
25. Bartheld, R.G., Habetler, T.G., Kamran, F.: Motor bearing damage detection using stator current monitoring. *IEEE Trans. Ind. Appl.* **31**(6), 1274–1279 (1995)
26. Index of /kat/BearingDataCenter: <https://groups.uni-paderborn.de/kat/BearingDataCenter/> (2016). Accessed 02 March 2019
27. Samanta, B., Al-Balushi, K.R., Al-Araimi, S.A.: Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Eng. Appl. Artif. Intell.* **16**(7–8), 657–665 (2003)
28. Obaid, R.R., Habetler, T.G., Stack, J.R.: Stator current analysis for bearing damage detection in induction motors. In: *Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, no. August (2003)
29. Demuth, H.B., Beale, M.H., De Jess, O., Hagan, M.T.: Neural network design. Martin Hagan, Stillwater (2014)
30. Mathew, J.: The condition monitoring of rolling element bearings using vibration analysis. *J. Vib. Acoust.* **106**, 447–453 (1984)

31. Prabhakar, S., Mohanty, A.R., Sekhar, A.S.: Application of discrete wavelet transform for detection of ball bearing race faults. *Tribol. Int.* **35**(12), 793–800 (2002)
32. Ocak, H., Loparo, K.A., Discenzo, F.M.: Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics. *J. Sound Vib.* **302**(4–5), 951–961 (2007)
33. Kimotho, J.K., Sextro, W.: An approach for feature extraction and selection from non-trending data for machinery prognosis. In: *Proceedings of the Second European Conference of the Prognostics and Health Management Society*, pp. 1–8 (2014)
34. Samanta, B.: Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mech. Syst. Signal Process.* **18**(3), 625–644 (2004)
35. Genetic Algorithm Options—MATLAB & Simulink: <https://www.mathworks.com/help/gads/genetic-algorithm-options.html> (2019). Accessed 2 March 2019
36. Umbarkar, A., Sheth, P.: Crossover operators in genetic algorithms: a review. *ICTACT J Soft Comput* (2015). <https://doi.org/10.21917/ijsc.2015.0150>
37. Deep, K., Pratap, K., Kansal, M.L., Mohan, C.: A real coded genetic algorithm for solving integer and mixed integer optimization problems. *Appl. Math. Comput.* **212**(2), 505–518 (2009)
38. Michalewicz, Z.: *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd edn. Springer, Berlin (1996)
39. Wang, B.: Data fused motor fault identification based on adversarial auto-encoder. In: *2019 IEEE 10th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, pp. 299–305 (2019)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.