Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 00 (2020) 000–000

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

30th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2021)
15-18 June 2021, Athens, Greece.

Intelligent fault diagnosis of rotating machine elements using machine learning through optimal features extraction and selection

Syed Muhammad Tayyab[a] *, Eram Asghar[b], Paolo Pennacchi [c], Steven Chatterton [d]

[a,b,c,d] *Dipartimento di Meccanica, Politecnico di Milano,Milan, Italy*

* Corresponding author. Tel.: +39-349-500-4501; E-mail *address:* syedmuhammad.tayyab@polimi.it

Abstract

The rolling element bearings, and gears are the main components of rotating machines and are most prone to defects which may result in significant economic loss. The main purpose of this study is an automated diagnosis of rolling element bearings and gears defects using machine learning (ML) technique and statistical features extracted from time domain vibration signal and spectral kurtosis. Extracted features are used to train K- nearest neighbors (KNN) as diagnostic classifier. The significance of segmentation size for time domain raw vibrational signals for the purpose of feature extraction is studied. This analysis is carried out by varying the window/segment length for features extraction and observing its effect on classification accuracy. Importance of feature selection for optimal performance of KNN in defect classification is studied by selecting most important and useful features using Genetic Algorithm (GA). Furthermore, effect of value of K on performance of KNN classifier has been observed by varying the value of K between 1 to 10 with step size of 1. Results show the ability of KNN classifier in combination with GA for correct and confident fault diagnosis of rotating machine elements in case of proper selection of parameters for features extraction.

Keywords: K-nearest Neighbor (KNN); Rolling element bearings and gears; Fault Diagnosis; Genetic Algorithm (GA)

1. Introduction

Todays' industry uses increasingly complex machines, with high demanding performance criteria. Early forecast of the anomalies in the equipment helps to carry out timely maintenance of component degradation. Therefore, performing effectual condition monitoring draws significant advantages to manufacturing industry [1,2]. Most of the time, the human intervention cannot respond efficiently and effectively to immense measurement data. Therefore, in order to minimize the human dependency, intelligent condition monitoring systems are required to be developed [3]. During last two decades, a lot of new approaches in this field have been proposed by researchers, especially the ones related to Artificial Intelligence (AI) techniques. Most of the approaches are applied on rotating machine elements focusing on bearing and gearbox defects. An approach was proposed for fault detection of rolling element bearings using neural network (NN) [4,5]. Principal component analysis (PCA) based feature selection scheme was presented for machine defect classification [6]. Another approach was proposed by Su H and Chong KT [7] for machine condition monitoring using support vector machine (SVM). Moreover, Liu H et al. [8] presented deep learning technique for fault detection using a rapid Fourier transform (STFT) scheme. A hierarchical diagnosis network (HDN) for fault pattern recognition was introduced by Gan M and Wang C [9]. Feature extraction method for the diagnosis of bearings using unsupervised data was proposed by Oh H et al. [10]. Another hierarchical approach for bearing defect diagnosis using adaptive deep convolutional neural network (CNN) was presented by Guo X et al. [11]. Zhang Z et al. [12] proposed an intelligent diagnosis and prognosis approach using integration of wavelet transform and PCA. For gear box defect diagnostics: Li C et al. [13,14] proposed multimodal deep support vector classification (MDSVC) and a deep random forest fusion (DRFF) technique, Chen Z et al. [15,16] used multiple layered neural network scheme and convolutional neural network method.

To improve the accuracy and efficiency of fault diagnosis of rotating machinery, application of Machine Learning techniques in machine condition monitoring still need more encouragement and attention. Therefore, the goal of this research is to develop an autonomous fault classification system, using appropriate artificial Intelligence technique, that can be used to improve the efficiency and applicability of condition monitoring technology.

The remainder of the paper is arranged as follows: in the next section recent literature related to fault diagnostics through AI techniques has been illustrated. Later, Section 3 consists of a case study in which the proposed approach is applied to the vibrational samples of two types of rotating machine elements separately. One is rolling element bearing and the other component is gear. In section 4, detailed analysis of the results obtained from the proposed approach has been carried out. The whole research work is being summarized in the section 5 along with some future recommendations.

2. Literature Review

Accurate fault detection and diagnosis systems have gained significant importance so that the potential failures of machinery can be managed properly. Various methods have been applied by researchers to tackle these issues. However, for condition monitoring of rotating machine elements, vibration monitoring is most widely used, as was published in [17]– [20]. Since the raw vibrational data is 1-dimensional time domain series data; extracting the appropriate features as health indicator is required. Those mechanical components that are often highly loaded in rotating machinery are gears and bearings (rolling element bearings in particular), responsible for early damages in machine's life. Rolling element bearing and gears are vital part of rotating machinery and their failure cannot only cause production losses but also lead to the financial losses. It is important to mention here that only bearing related defects are responsible for more than 40% of the failure in industrial machines [21]. In this instance, a lot of research in the field of fault diagnosis has been carried out since last few decades. Qicai Zhou et al. presented a study in which K-mean algorithm was used to label the un-labeled signals [18]. A comparison of the performance of ANN and KNN was carried out by Rohit S. Gunerkar et al. [19]. The study was performed on rolling element bearings based upon feature extracted after wavelet transform. J. P. Patel did comparison between ANN and SVM considering bearings and found that SVM gave better results as compared to ANN [20]. Few years back, a method based on spectral kurtosis (SK) and cross correlation was presented by Jing Tian et al [22] to diagnose defects and monitor the degradation of bearings in electric motors. The method was validated by experiments using machinery fault simulator. Daniel Augusto [23] carried out analysis to diagnose faults in water cooling system using data collected from real engine at all speeds to train the classifier. The performance of KNN and ANN classifiers was compared and found same using average mean square and classification error.

Many studies have been made in past few decades on the fault detection of mechanical components using AI techniques [24,25]. Since using these techniques for diagnosing defects have various benefits if compared with the traditional mathematical and statistical modelling approaches. However, there are lots of things still need to be improved for these methods in order to make it more effective and practical to solve real world applications [26]. Furthermore, Genetic Algorithm (GA) applications with vibrational signal analysis in machine condition monitoring and fault diagnosis still need more support and attention because of the lack of existing evidence. Vibration based conditional monitoring is most popular for early monitoring and identification of defects. However, localized faults produce very weak impulses in vibrational signals. Therefore, it is difficult to detect these faults using existing frequency domain methods [27]. Various supervised machine learning methods have so far been used for fault identification of rolling element bearings and gears using time domain statistical features. However, the fault classification through the usage of minimum features is still a challenge for researchers.

In the light of existing literature, it was observed that there is still need to improve the quality of the features extracted from vibrational data so that they may contain appropriate and enough information regarding machine health in order to properly train ML classifier for defect diagnosis. In this research work, vibration-based monitoring technique is used for fault diagnostics of rolling element bearings and gears. Statistical features derived from time-domain signal and spectral kurtosis are extracted. Then, K- Nearest Neighbor is applied upon the combination of all extracted features to classify the faults. The performance of classifier is compared for different values of K in order to determine its potential in classifying various faults of bearings and gears. Afterwards, Genetic algorithm is used to select most suitable features for optimizing the performance of KNN classifier.

3. Case Study

For accurate mechanical fault diagnosis realistic and correct features extraction from vibration signal; containing enough information regarding machine/component's health is of paramount importance. For feature extraction from vibration data, it is to be divided into samples having same number of points (dimension). The number of points in one sample will be referred as window size (W) for feature extraction in this paper. In order to find effect of window size for feature extraction on classification accuracy, it is essential to set some reference. For this purpose, window size corresponding to one rotation (Wr) is considered as reference in this paper, which can be described as:

$$Wr = 60\,Fs/R \tag{1}$$

Where; Fs is the sampling frequency of acceleration sensor in Hz and R is the revolution per minute of the rotating machine. The case study was carried out for both components (rolling element bearings and gears) separately. Statistical features derived from time-domain signal and spectral kurtosis (SK) are extracted. Initially, 16 features (root mean square, peak to peak, kurtosis, crest indicator, impulse, energy-I, skewness, standard deviation, variance, shape factor, margin factor, energy-II, spectral kurtosis means, SK kurtosis, spectral kurtosis skewness, spectral kurtosis standard deviation) were calculated keeping the window size W = 400 ($\sim Wr/4$ for bearings' and $\sim Wr$ for gears' data set). The dimension of feature matrix for rolling element bearing and gear data set was 609x16x4 and 250x16x3, respectively. In order to analyze the effect of window size W for feature extraction on classifier accuracy, it was increased to 800 ($\sim Wr/2$) and 1200 ($\sim 2Wr/3$) for rolling element bearing and to 800 ($\sim 2\,Wr$) for gears.

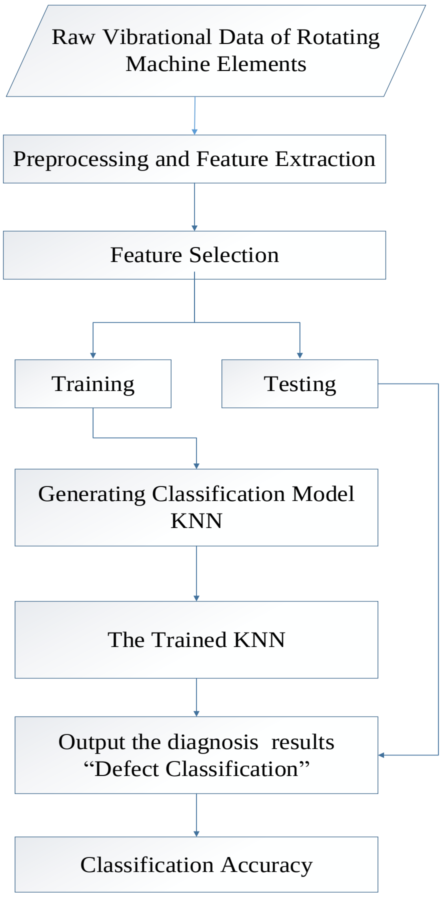


Figure. 1: Working Methodology

Since the requirement of this era is to automatize the fault diagnosis to avoid human intervention in this tedious task. For this instance, the classification technique from machine learning (K- Nearest Neighbors) has been applied to diagnose the defects of rotating machine elements. The flow of the adopted methodology is illustrated in Fig.1.

3.1. Experimental setup and vibrational data

The rolling element bearing vibration signals used for analysis were downloaded from CWRU Bearing Centre [28] as shown in Fig.2. Vibration data are collected for motor loads from 0 to 3 hp and motor speeds from 1,720 to 1,797 rpm using two accelerometers installed at both the drive end and fan end of the motor housing, and two sampling frequencies of 12 kHz and 48 kHz were used. Four classes of data with 48 kHz sampling frequency were selected for research as mentioned in Table-1. Experimental data for gear fault diagnosis was acquired from online shared gear faults data sets [29] as shown in Fig. 3. In each case of gears data set, the acceleration was recorded for a duration of 10 second with the sampling frequency of 10 kHz. The gearbox fault data was collected in three different pinion conditions (see Table-2).

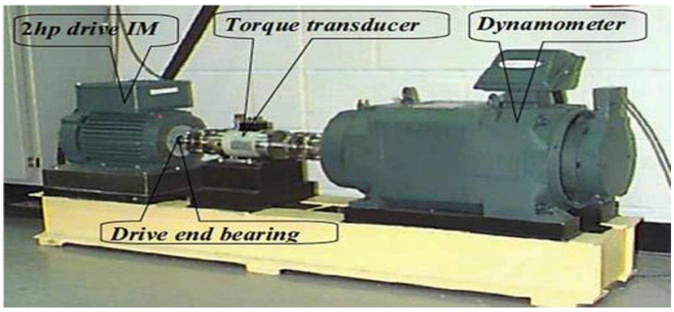


Figure. 2: Experimental test setup for bearing data set [28]

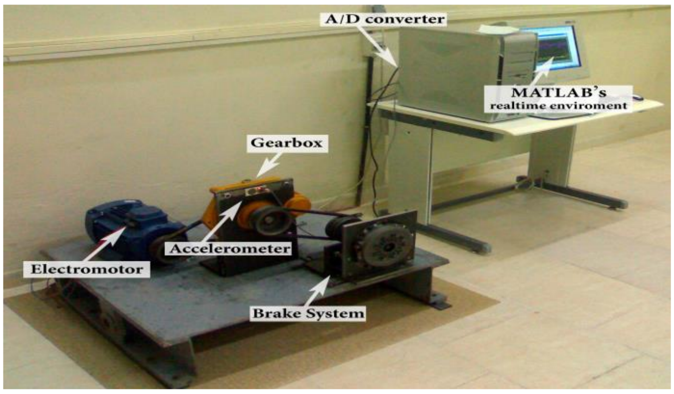


Figure. 3: Experimental test setup for gear data set [29]

Table 1. Selected bearing fault classes

Classes	Type of defect	Fault diameter	Motor Load (HP)	Motor Speed (rpm)
1	Normal	No fault	0	1797
2	Ball	0.021"	0	1797
3	Outer Race Centered @6:00	0.021"	0	1797
4	Inner Race	0.021"	0	1797

Table 2. Selected gear fault classes

Classes	1	2	3
Type of Faults	No fault	One chipped tooth	Three consequent worn teeth

3.2. Application of K-nearest neighbors algorithm

KNN is an algorithm that learns from all available patterns (training data) and classifies the new patterns (testing data) based on the similarity measure between them. The similarity measure is the minimum distance (e.g. Mahalanobis distance, Euclidean distance etc.) between the data points. It requires to set the number of nearest neighbors (k) that are considered for classification based upon selected similarity measure. The classification through KNN algorithm is divided into training and testing phases. The available data is divided into training and testing samples. During the training phase, KNN model finds the relationship between predictors and targets of training data whereas; in testing phase, trained model is tested for its ability to predict the correct classes of testing data whose labels are known but the model has not already been trained on this data.

In this research, analysis is carried out to ascertain the effect of K values and window size for feature extraction to improve the classification accuracy of model. Euclidean distance is used in this study to find the similarity between the data points which is simple to implement and can give good results. Lesser the distance between data points, higher is the similarity. Initially, KNN model is trained and tested using all extracted features. However, in order to reduce the dimensionality of feature matrix for the purpose of reducing the computational requirements and improving the classification accuracy, optimal features are selected through wrapper approach. Genetic Algorithm (GA) is used for effective feature selection for isolating the best features evaluated through fitness criteria. The significance of using GA over other methods is that it consumes less computational power i.e. no need for derivative calculations. It is more robust algorithm and has a superior global searching capability in complicated search spaces. On contrary, GA needs preliminary tuning and typically many fitness evaluations are required to get better solutions. In this case study, GA parameters have been defined carefully and algorithm was run for enough number of generations to achieve effective convergence. The fitness/ objective function of this algorithm returns the numeric value calculated against every sample for randomly selected features. This selection is controlled through the values generated in the binary encoded genome. The performance of KNN classifier in terms of minimizing the classification error by resubstitution/ resubstitution loss is used as the fitness function for GA.

3.3. Optimization through Genetic Algorithm

GA is used for selection of best features in order to reduce the dimensionality of features matrix and to optimize the performance of KNN. The algorithm starts by initializing the population (design variables/ labelled data) which is later evaluated based on considered fitness function. This population of variables is user defined and needs to be set appropriately. GA can be used for feature selection using binary or real coded individuals. In this case study the population of individuals is randomly generated “genomes” which is a binary encoded bit. The reason behind using the binary bits is that it helps to reduce the likelihood of convergence within the population. The whole subset of variables is subjected to three GA operators; selection, crossover and mutation. Genetic operators are based on the idea of heredity of genetic characters over the generations. By the application of these operators, best solutions/individuals are separated resulting in the convergence of the fitness towards global best solution over the generations till the stopping criteria meets [30]. Most common fitness evaluation or selection methods are proportionate, ranking and tournament selection. In proportionate selection method individual fitness is divided by the average fitness of population; roulette wheel is one of the most used method in which slice of circular roulette wheel is assigned to individuals. Upon spinning it up to the “number of population”; gives the parents and child sequences which is repeated iteratively. In the second method fitness scale is not required because every chromosome is ranked based on its own fitness, this helps in preventing the early convergence. The disadvantage is that the ratio of expected values of individuals remains same even when the fitness variance is low or high. In the tournament selection two individuals are randomly selected from initial population and the best individual from those two is preserved for next generation. In GA, selection method is of critical importance, it aids in finding out the individuals which need to be prioritized to produce next generation (child chromosomes). Thus, setting it carefully may lead the solutions towards desired optimality level. If the selection is strong and strict, highly fit suboptimal solution may take over population. On contrary, slow convergence is normally observed in case of weak selection. In this case study, tournament selection has been used as selection criteria; two individuals executed from the initialized population are set out for a tournament and the best individual is preferred as a parent. Most discriminating feature of GA over other search methods is crossover as it contributes in maintaining diversity by introduction of new genetic content when operated on two chromosomes to produce off-springs. In order to accelerate the search in population recombination operator is applied. Simplest way of crossover is to choose a cutoff point randomly and exchange the genetic content by combination of single segment of two parents to give child sequence. There are many methods to accomplish recombination. To achieve better performance and better combinations, properly designed crossover mechanism is required thus recombination should be carried out with some probability called crossover rate. Crossover rate is the ratio of number of child sequences produced in each generation to the population size. The third GA operator is mutation; it simply alters one or more genes that are lost during selection operator to maintain diversity. Although it is secondary operator but its probability of applying is as important as that of crossover. Most common mutation is the bit flip mutation in which it randomly selects two individuals and swaps them with a probability called mutation rate [30]. Moreover, in the applied algorithm of GA, diversity of the solutions has been maintained by preserving the best individuals for the next generations, this phenomenon is called elitism. The algorithm for GA has been designed in MATLAB using the input extracted from the features (labelled data). GA rely on its population unlike traditional techniques. Thus, population size is user defined and it has drastic effects on the performance of GA. Small population size may result into premature convergence and large population size takes unnecessary computational time. In this case study population size is set to be 50. The algorithm for GA is represented in Fig 4. Initializing the population and then converting the individuals into binary encoding facilitates in finding out the fitness of all individuals based on the selection fitness/evaluation criteria. GA does not have intense mathematical requirements, so it has the tendency to handle any type of fitness/ objective function and constraints. In this case study, the objective function of GA is to minimize the classification error. In every generation it selects the individuals having minimum classification error and use these individuals for next generations and discard the other ones considering equal importance to the data of all classes. The best individuals are the offspring for current generation and parents for the next successive iteration, evolving the cycle till the stopping criteria is fulfilled.

4. Results and Discussion

Performance of KNN classifier in terms of classification accuracy for both rotating machine elements (rolling element bearings and gears) is appended at Table 3. For rolling elements bearings three window sizes ($W1$= 400≈ $Wr/4$, $W2$= 800 ≈ $Wr/2$ and $W3$ =1200 ≈ $2Wr/3$) were used for feature extraction. All 16 extracted features were used for training the KNN model. For W=400, classification accuracy was observed more than 90% for all values of k under observation. The minimum accuracy was 90.3% for K=2 and maximum accuracy was 93.6% for K= 5&7. Confusion matrix for K=4 at this window size is shown in Fig. 5a. By increasing the window size to 800, classification accuracy was improved for all values of K under consideration. For $W = 800$, minimum classification accuracy of 97.3% was observed for K=2 and maximum accuracy of 98.2 % was observed for K=4. Confusion matrix for K=4 at this window size is shown in Fig 5b. Afterwards, window size was further increased to 1200. For this window size, 99.6% classification accuracy was observed for K= 2 to k=10. Classification accuracy at this window size ($W = 1200$) was further improved to 100% after applying Genetic Algorithm (GA) for features selection. GA selected only 04 features (Root Means Square, Impulse, Kurtosis and Shape Factor) which reduced the dimensions of features matrix. Confusion matrix for K=4 at this window size are shown in Fig 6a and 6b.

For gears data set two window sizes ($W1 = 400 \approx Wr, W2 = 800 \approx 2Wr$) were used for feature extraction. Firstly, all 16 features were used for the classification. For window size = 400, classification accuracy of above 92% was observed for all values of K between 1-10. Maximum accuracy was observed to be 94.6% at K=4. Confusion matrix for K=4 at this window size is shown in Fig 7. By increasing the window size to 800 classification accuracy was significantly improved for all values of K. Minimum classification accuracy was observed to be 98.2% for K=8 and 100% classification accuracy was achieved for K=2-5 (see Fig. 8a). When GA was applied for feature selection, 03 features (kurtosis, skewness and standard deviation) were selected. Accuracy of classification was improved to 100% for K=2-9 (see Fig 8b). Dimensionality of features matrix was reduced. For both data sets it is observed that performance of KNN in terms of classification accuracy is not much sensitive to value of K between 1-10. However, by increasing the window size for features extraction to 800,1200 for bearings and to 800 for gears, performance of both classifiers was remarkably improved. It shows that for proper feature extraction which will carry the correct information regarding the equipment defect, it is essential to use appropriate size of window. It is observed that acceptable classification accuracy can be achieved by setting $W \geq Wr/2$ for rolling element bearings and $W \geq 2Wr$ for gears. Features extracted by using smaller windows will not carry adequate information for correct fault classification.

Feature selection through GA was very useful in order to find most important features for fault diagnostics. Individuals having the best fitness were obtained for generation size 100 (stopping criteria) and population size of 50. Crossover and mutation operator need to be designed appropriately with some probability rate to attain better performance. In this case study the crossover between two samples is being carried out at a rate of 0.8 and the same rate has been used for mutation. KNN did fault classification with 100% accuracy using only 04 and 03 features selected through GA in case of bearings and gears, respectively. Furthermore, the performance of classifier was improved in terms of computational time as well. The computational time was reduced up to 73% in case of bearings and up to 48% in case of gears data set. Figures 9 (a) and (b) show reduction in computational time because of dimensionality reduction for bearings and gears data sets, respectively. Performance comparison of this study with few existing studies, who have used the Case Western Reserve University Bearing data set, is given in Table 4.

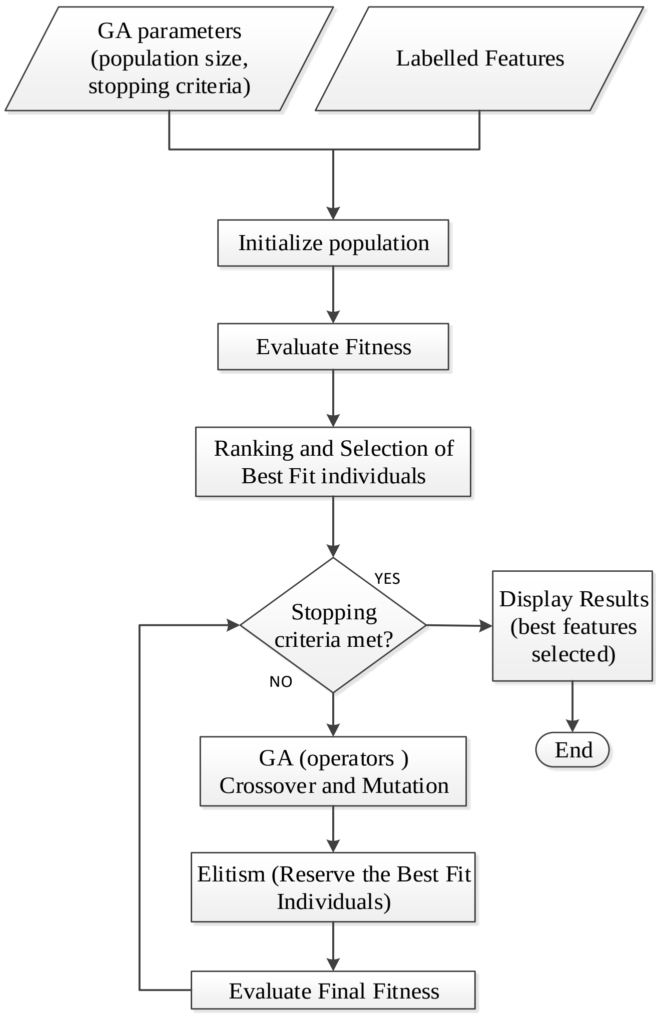
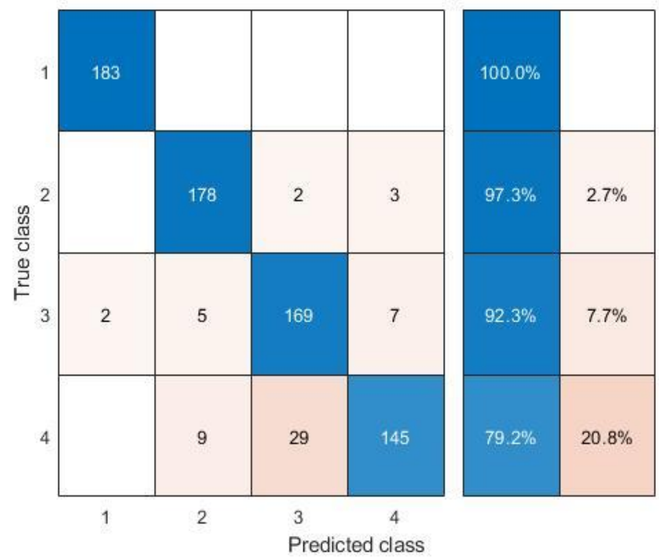


Figure. 4: Genetic algorithm for feature selection

Table 3. Classification Accuracy and Computational Time

Rolling Element Bearings								Gears					
ML Model	**KNN**												
Window Size	**400**	**800**	**1200**					**400**	**800**				
Number of Features	16	16	16	4(GA)	16	4(GA)	**Reduction of Time (%)**	16	16	3(GA)	16	3(GA)	**Reduction of Time (%)**
K	**Classification Accuracy (%)**				**CPU Time (Sec)**			**Classification Accuracy (%)**			**CPU Time (Sec)**		
1	90.7	98.0	99.2	100	4.52	2.24	50	92.0	99.1	99.1	2.78	1.69	39
2	90.3	97.3	99.6	100	4.09	2.12	48	93.7	100.0	100	2.68	1.61	40
3	93.4	97.7	99.6	100	4.47	2.28	49	93.8	100.0	100	2.70	1.67	38
4	92.2	98.2	99.6	100	4.05	1.84	54	94.6	100.0	100	2.77	1.80	35
5	93.6	97.7	99.6	100	4.64	1.86	60	94.2	100.0	100	2.77	1.77	36
6	92.9	97.7	99.6	100	5.60	1.98	65	92.0	99.1	100	2.58	1.66	36
7	93.6	97.7	99.6	100	5.31	2.02	62	93.7	99.1	100	2.70	1.61	40
8	93.2	97.7	99.6	100	5.97	1.59	73	92.9	98.2	100	2.64	1.38	48
9	93.2	97.7	99.6	99.6	4.83	1.66	66	92.9	99.1	100	2.61	1.36	48
10	93.3	98.0	99.6	99.6	4.58	1.75	62	92.0	98.3	99.1	2.70	1.85	32

a



b

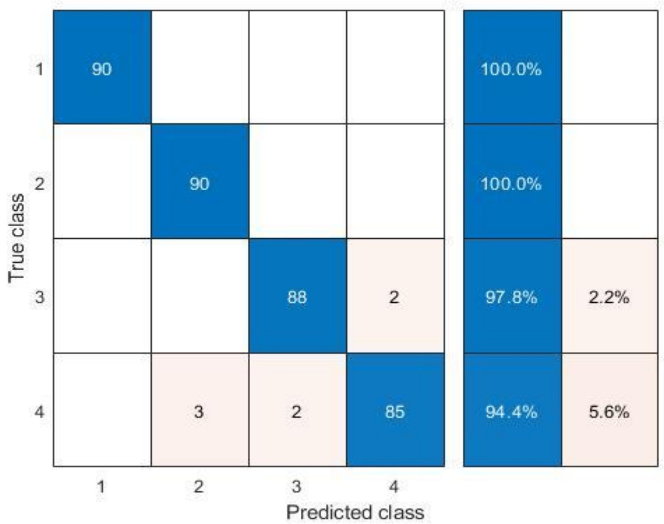
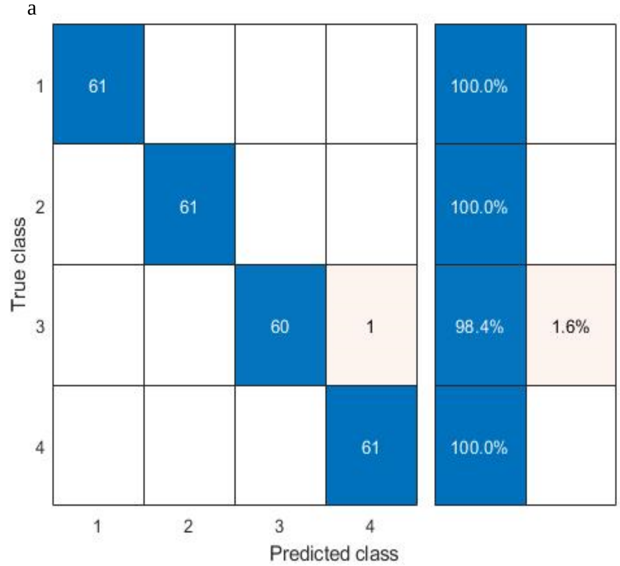


Figure 5: Confusion charts for rolling elements bearings for (a) window size 400; (b) window size 800

a



b

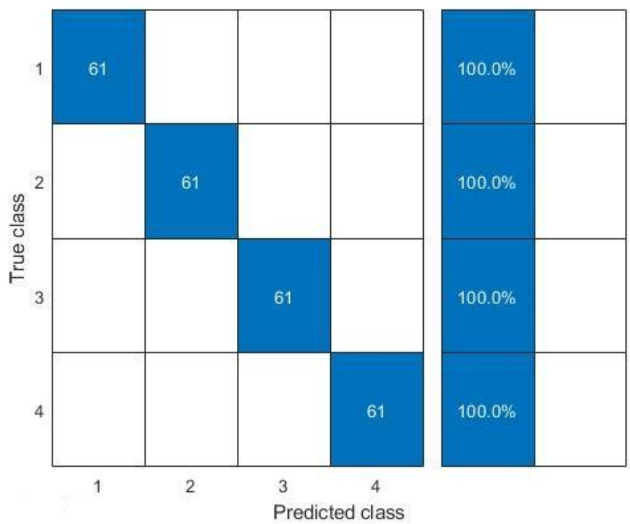


Figure 6: Confusion charts for rolling elements bearings for window size 1200 (a) all features; (b) with selected features

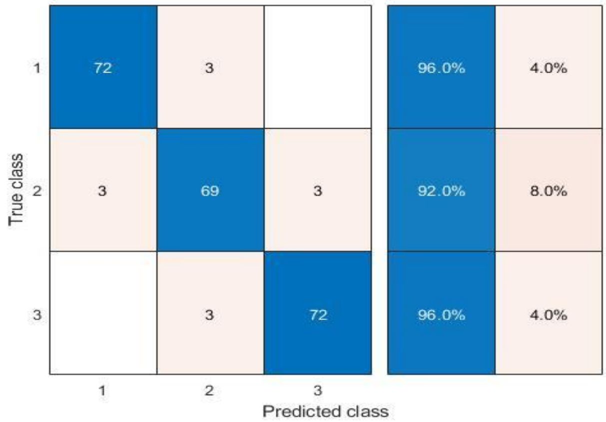


Figure 7: Confusion chart for gears at window size 400

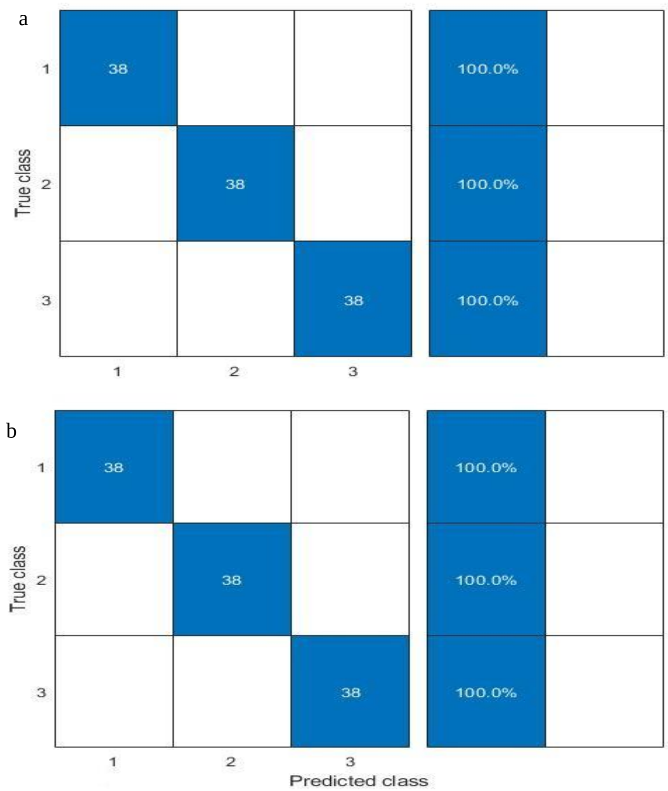


Figure 8: Confusion chart for gears at window size 800 (a) considering all features; (b) selected features

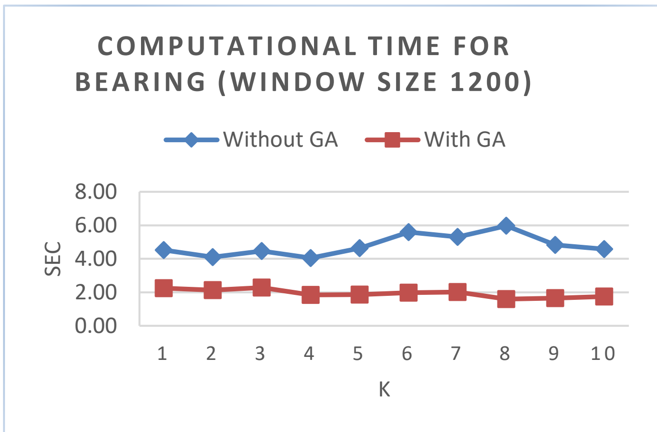


Figure 9a: Computational time comparison for bearings

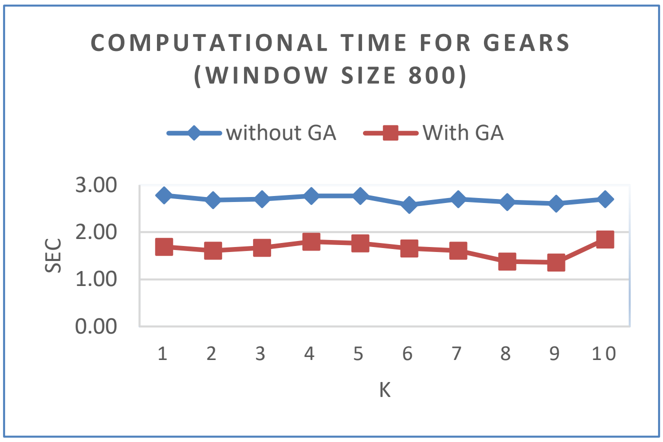


Figure 9b: Computational time comparison for gears

Table 4. Comparison with existing approaches

Work /Approaches	Extracted Features	Feature Selection Approach	Features after Reduction	Types of Bearings used	Considered Classes	Classifier	Misclassification Rate
Zhang etal.,2013 [31]	21	Kernel Principal Component Analysis	3	7	Drive End	SVM	0.47%
Jiang etal.,2013 [32]	16	Semi-supervised kernel Marginal Fisher Analysis	5	10	Drive End	SVMKNN	0.00% 1.50%
Li etal.,2013 [33]	14	Linear Local Tangent Space Alignment	3	7	Drive End	Littlewoods-Paley SVM	5.71%
Zhu etal.,2014 [34]	8	None	8	10	Drive End	SVM	0.00%
Baraldi etal.,2016 [21]	29	Wrapper search	5	10	Drive End	KNN	0.01%
Ours	16	Genetic Algorithm	4	4	Drive End	KNN	0.00 %

Conclusion

A case study was conducted for defect diagnosis of rotating machine elements (bearings and gears) using KNN classifier and combination of statistical features extracted from time-domain signal and spectral kurtosis. Importance of appropriate segmentation of time domain raw vibrational signal for feature extraction and subsequent feature selection through GA for defect classification was also studied. Different conditions for rolling element bearings and gears were considered and observed that feature extraction has significant effect on diagnosis results. Selecting very small size of segment/window for feature extraction to increase the amount of training data, adversely affects the features quality and subsequently affects the classification accuracy of ML classifier. Feature selection was not much useful for performance improvement of KNN classifier in this case, because the classifier had already shown very good performance while using all 16 features. However, dimensionality of feature matrix was remarkably reduced, improving the classifier performance in terms of less computational time, thus making the demonstrated method suitable for online monitoring and development of compact autonomous fault diagnosis systems as well. In future, the comparison of GA will be carried out with other feature selection and dimensionality reduction methods using recommended optimal window size.

References

[1] Altmann J. *The Application of Discrete Wavelet Packet Analysis for the Detection and Diagnosis of Low Speed Rolling-element Bearing Faults* (Doctoral dissertation, Monash University).

[2] Baillie DC, Mathew J. A comparison of autoregressive modelling techniques for fault diagnosis of rolling element bearings. Mechanical Systems and Signal Processing. 1996 Jan 1;10(1):1-7.

[3] Shebin EL. Fault diagnosis of rotating machinery based on vibration analysis. International Journal of Advanced Engineering and Global Technology I. 2016;4(1).

[4] Li B, Chow MY, Tipsuwan Y, Hung JC. Neural-network-based motor rolling bearing fault diagnosis. IEEE transactions on industrial electronics. 2000 Oct;47(5):1060-9.

[5] Samanta B, Al-Balushi KR. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. Mechanical systems and signal processing. 2003 Mar 1;17(2):317-28.

[6] Malhi A, Gao RX. PCA-based feature selection scheme for machine defect classification. IEEE Transactions on Instrumentation and Measurement. 2004 Nov 22;53(6):1517-25.

[7] Su H, Chong KT. Induction machine condition monitorin using neural network modeling. IEEE Transactions on Industrial Electronics. 2007 Feb 5;54(1):241-9.

[8] Liu H, Li L, Ma J. Rolling bearing fault diagnosis based on STFT-deep learning and sound signals. Shock and Vibration. 2016;2016.

[9] Gan M, Wang C. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. Mechanical Systems and Signal Processing. 2016 May 1;72:92-104.

[10] Oh H, Jeon BC, Jung JH, Youn BD. Smart diagnosis of journal bearing rotor systems: Unsupervised feature extraction scheme by deep learning. InProceedings of the Annual Conference of the Prognostics and Health Management Society, Denver, CO, UAS 2016 Oct (pp. 3-6).

[11] Guo X, Chen L, Shen C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement. 2016 Nov 1;93:490-502.

[12] Zhang Z, Wang Y, Wang K. Intelligent fault diagnosis and prognosis approach for rotating machinery integrating wavelet transform, principal component analysis, and artificial neural networks. The international journal of advanced manufacturing technology. 2013 Sep 1;68(1-4):763-73.

[13] Li C, Sanchez RV, Zurita G, Cerrada M, Cabrera D, Vásquez RE. Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. Neurocomputing. 2015 Nov 30;168:119-27.

[14] Li C, Sanchez RV, Zurita G, Cerrada M, Cabrera D, Vásquez RE. Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. Mechanical Systems and Signal Processing. 2016 Aug 1;76:283-93.

[15] Chen Z, Li C, Sanchez RV. Gearbox fault identification and classification with convolutional neural networks. Shock and Vibration. 2015;2015.

[16] Chen Z, Li C, Sánchez RV. Multi-layer neural network with deep belief network for gearbox fault diagnosis. Journal of Vibroengineering. 2015 Aug 15;17(5):2379-92.

[17] Kiral Z, Karagülle H. Simulation and analysis of vibration signals generated by rolling element bearing with defects. Tribology International. 2003 Sep 1;36(9):667-78.

[18] Orhan S, Aktürk N, Celik V. Vibration monitoring for defect diagnosis of rolling element bearings as a predictive maintenance tool: Comprehensive case studies. Ndt & E International. 2006 Jun 1;39(4):293-8.

[19] Wu JD, Hsu CC. Fault gear identification using vibration signal with discrete wavelet transform technique and fuzzy–logic inference. Expert systems with applications. 2009 Mar 1;36(2):3785-94.

[20] Vilakazi CB. *Machine condition monitoring using artificial intelligence: the incremental learning and multi-agent system approach* (Doctoral dissertation).

[21] Baraldi P, Cannarile F, Di Maio F, Zio E. Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions. Engineering Applications of Artificial Intelligence. 2016 Nov 1;56:1-3.

[22]Tian J, Morillo C, Azarian MH, Pecht M. Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis. IEEE Transactions on Industrial Electronics. 2015 Dec 17;63(3):1793-803.

[23] Do Carmo DA, Munaro CJ. Application of kNN and artificial neural network classifiers for fault detection and diagnosis in a diesel engine cooling system. In2018 13th IEEE International Conference on Industry Applications (INDUSCON) 2018 Nov 12 (pp. 1067-1071). IEEE.

[24] Aherwar A. An investigation on gearbox fault detection using vibration analysis techniques: A review. Australian Journal of Mechanical Engineering. 2012 Jan 1;10(2):169-83.

[25] Chang CW, Lee HW, Liu CH. A review of artificial intelligence algorithms used for smart machine tools. Inventions. 2018 Sep;3(3):41.

[26] Ali YH. Artificial Intelligence Application in Machine Condition Monitoring and Fault Diagnosis. Artificial Intelligence: Emerging Trends and Applications. 2018 Jun 27:275.

[27] Wowk, Victor. "Machinery Vibration MeaSurement and Analysis, McGraW-Hil." (1999)

[28] Case Western Reserve University Bearing Data Center Website. Available online:https://csegroups.case.edu/bearingdatacenter/pages/download-data-file (accessed on 18 December 2019).

[29] Experimental Dataset for Gear Fault Diagnosis", Amir Hosein Zamanian, Southern Methodist University, Dallas, TX, 75205, USA. Available online: https://goo.gl/TorZJq (accessed on 18 December 2019

[30] Mitchell M. An introduction to genetic algorithms. MIT press; 1998.

[31] Zhang, Y.,Zuo,H.,Bai,F.,2013.Classification of fault location and performance degradation of a roller bearing. Meas:J.Int.Meas.Confed.46(3),1178–1189.

[32] Jiang, L., Xuan, J., & Shi, T. (2013). Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis. *Mechanical Systems and Signal Processing*, *41*(1-2), 113-126.

[33] Li, F.,Tang,B.,Yang,R.,2013.Rotating machine fault diagnosis using dimension reduction with linear local tangent space alignment. Meas.: J. Int. Meas. Confed. 46 (8),2525–2539.

[34] Zhu, K.,Song,X.,Xue,D.,2014. A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm. Measurement 47, 669–675.