

# Significance of Incorporating Chebfunction Coefficients for Improved Machine Fault Diagnosis

**Sachin Saj T K, Neethu Mohan, Sachin Kumar S, Soman K.P**

Center for Computational Engineering & Networking (CEN), Amrita School of Engineering,  
Coimbatore, Amrita Vishwa Vidyapeetham, India.

E-mail: [sachin96saj@gmail.com](mailto:sachin96saj@gmail.com)

**Abstract.** For any industry the efficiency and performance of rotating machinery/mechanical systems is a major concern. Bearings and gears are two essential parts in a rotating machinery and any defects in these components can lead to a major breakdown of the system thus causing large economical loss for the company. An appropriate machine condition monitoring system is essential in such scenarios for identifying the health of the machines. Therefore in this paper fault diagnosis of rotating mechanical systems is performed as a feature dependent-pattern classification problem. The machine is made to run in different good as well faulty conditions and the vibration signals are collected. Then chebfunction coefficients are extracted from the vibration signals as part of the feature extraction process. Finally, the extracted features are classified using regularized least squares (RLS) for identifying the good and faulty bearing as well as gear conditions of the machine. The evaluations are performed using different kernel functions and the average accuracy reported is 98% for bearing and gear data. The various experiments performed claims that the proposed system can be used for real-time fault diagnosis in rotating mechanical systems with sufficient accuracy.

## 1. Introduction

From last two decades there is an extensive increase in manufacturing industries since the manufacturing capabilities of the machines have increased exponentially [1, 2, 3]. Bearings and gears are two important parts in any rotating mechanical systems and faults in these parts can lead to complete breakdown of the system. This can cause a huge economic loss for the company. One of the primary failure mode of bearings and gears are localized defects such as cracks [4]. The detection and classification of the faults/defects become a crucial challenge in the modern scenario. Machine condition monitoring helps us in continuously monitor the machine parts such as bearing, gears etc. It gives indications of future failures, so that proper maintenance measures can be done prior itself [5, 6]. This will avoid huge break down of the machinery thus saving millions of rupees for the company. A machine condition monitoring system continuously collect the data, process the data and interpret the data [7]. The three main stages in machine fault diagnosis system are 1) signal acquisition-the collection of good/faulty data from rotating machines belongs to this stage, 2) feature extraction-this stage consists of the identification of suitable features and its extraction from collected data, 3) pattern classification- identification of normal and abnormal signals by feeding the extracted features to suitable classifier belongs to this final stage.

Vibration analysis is the most successful method that is used in machine condition monitoring

systems [8]. Every machines will generate certain kind of vibrations during their operation, even a good condition machine. The only difference is that the frequencies of vibrations produced by good conditioned machines and fault existing machines are different and this is the principle of machine condition monitoring system using vibration signals. Additional benefit of using vibration signals is that it contains vital information related to the internal forces of the system thus it can provide much more accurate results.

This paper proposes a feature dependent-machine condition monitoring system using chebfun coefficients and regularized least squares (RLS) for machine fault diagnosis. In the proposed system, the collected vibration signals from different good and faulty conditions are subjected to the feature extraction stage using chebfun approximation system. The suitable chebfun coefficients are taken as the features and are classified using RLS classifier. Various experiments using different kernel functions are performed to evaluate the effectiveness of the proposed fault diagnosis system. The remaining section of the paper is organized as follows: Section 2 gives a brief explanation about chebfun approximation system and RLS classifier. The experimental setup for vibration signal acquisition is discussed in Section 3 and Section 4 explains the two datasets used for experiments. Section 5 narrates the proposed system for machine fault diagnosis and Section 6 describes the major results of the proposed system. Finally conclusions are given in section 7.

## 2. Background

### 2.1. Chebfun Approximation Function

Chebfun (Cheb function) system is a recently developed platform that can be used for performing numerical computation with functions [9]. Chebfun approximation offers high speed and precise accuracy for computations. In chebfun approximation, functions are represented by chebyshev polynomial expansions. The chebyshev points are computed by preserving machine precision (roughly to machine precision is about 15-digit accuracy) [10]. The chebfun system deals with simple linear algebra problems to complicated non-linear problems. The example for constructing chebfun for a given function is illustrated as follows.

```
F = chebfun('sin(10*x)')
plot(F)
```

Figure 1 represents the chebfun approximation corresponds to  $\sin(10x)$  in continuous domain. From figure 1 it is clear that the default interval for chebfun approximation is  $[-1 \ 1]$ .

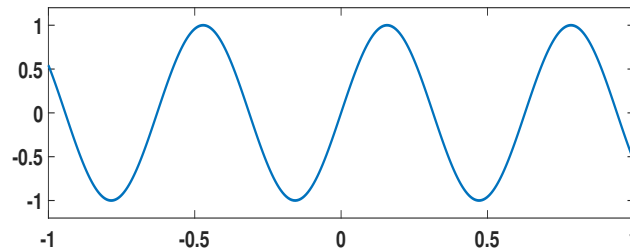


Figure 1: Chebfun illustration of  $\sin(10x)$

```
length(F)
Ans => 34
```

The length of chebfun object 'F' specifies that this function is represented by using 34 chebyshev points, that is the resulting chebyshev polynomial has a degree of 34. Figure 2 represents the

visualization of this 34 chebyshev points. For a given function the corresponding chebfun is a polynomial interpolant using chebyshev points. It is defined as in equation 1.

$$x_i = \cos\left(\frac{\pi i}{n}\right), 0 \leq i \leq n \quad (1)$$

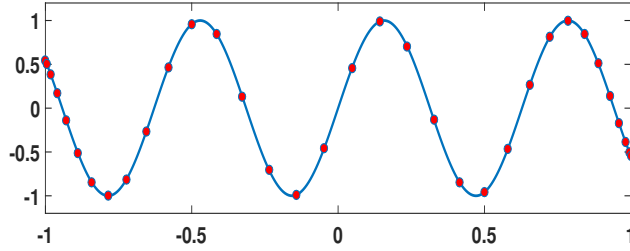


Figure 2: Representation of 34 chebyshev points

The distribution of points towards the end is more compared to the middle, and at -1 and 1 there are high number of points representing the signals. Recently, chebfun system is being used in domains such as control problems in aeronautics [11], power system [12], speech [13] etc.

### 2.2. Regularized Least Squares Classifier

Regularized least squares (RLS) classifier is one of the basic algorithm used for supervised learning [14]. In RLS, the training of the data is done by solving a system of linear equations. In this proposed method, the chebfun coefficients generated as our features is given to the RLS classifier. For an  $n$  dimensional data vector,  $X = \{x_1, x_2, \dots, x_n\}$  with corresponding labels  $Y = \{y_1, y_2, \dots, y_n\}$ , RLS predicts the output by computing the weight matrix (W) [2]. The objective function is defined as,

$$\min_{W \in R^{n \times T}} = \frac{1}{n} \|Y - WX\|_F^2 + \lambda \|W\|_F^2 \quad (2)$$

Where  $\lambda$  is the control parameter. The weight matrix can be finally obtained as,

$$W = (X^T X + \lambda I) X^T Y \quad (3)$$

The detail explanation related to the derivation of RLS algorithm is given in [15, 16].

## 3. Experimental Setup

In this paper we have used two different experimental setups: A) Bearing fault simulator system, B) Helical gear simulator system for generating the vibration signals.

### 3.1. Bearing Fault Simulator System

The experimental setup shown in figure 3 is the bearing fault simulator connected to sensors [5, 8]. This setup consists of variable DC motor of max 0.5 HP with a speed ranging from 0 to 3000 rpm. A shaft of diameter 30 mm is placed to motor through an aluminum flexible coupling. The reason for doing this is to minimize the effects caused by the misalignment and transmission of vibration from the motor. The setup consists of a control panel for the entire monitoring of the system. The details about this fault simulator setup is available in [2, 5, 8].

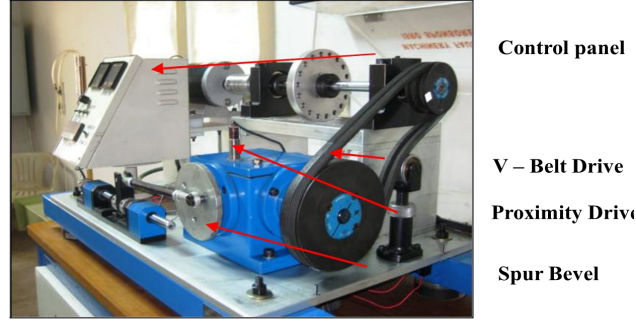


Figure 3: Experimental setup for bearing fault simulator

### 3.2. Helical Gear Fault Simulator System

Helical gear is one of the most important gear used for transmission as it produces large amount of thrust [17]. This fact motivated us to use gear fault system for condition monitoring. For the experiment, a two-stage helical gearbox of 5 HP is used. A 3-phase induction motor is used to drive this two-stage helical gearbox. An inverter drive is used to control the motor speed. The first stage gearbox speed is about 80 rpm and the speed of the smaller gear shaft is 1200 rpm. This shaft is attached to DC motor that generates about 2 KW power. Hence the final resultant load on the gearbox is 2.6 HP that is only 52% out of its rated power of 5 HP. The whole setup is mounted on I beams. The vibration signal coming out of the setup is collected through the accelerometers connected to the gearbox.

## 4. Vibration Data Acquisition

This section explains details about the vibration data generated using the experimental setup described in Section 3.

### 4.1. Bearing Fault Dataset (Dataset-1)

The bearing fault dataset (dataset-1) consists of 4 classes such as A: good condition, B: inner race fault condition (IRF), C: outer race fault condition (ORF), D: inner and outer race fault condition (IORF). Figure 4 represents the examples of signal corresponds to these 4 generated classes. The signals are generated with sampling frequency of 12000 Hz and sampling length of 8192 for all conditions.

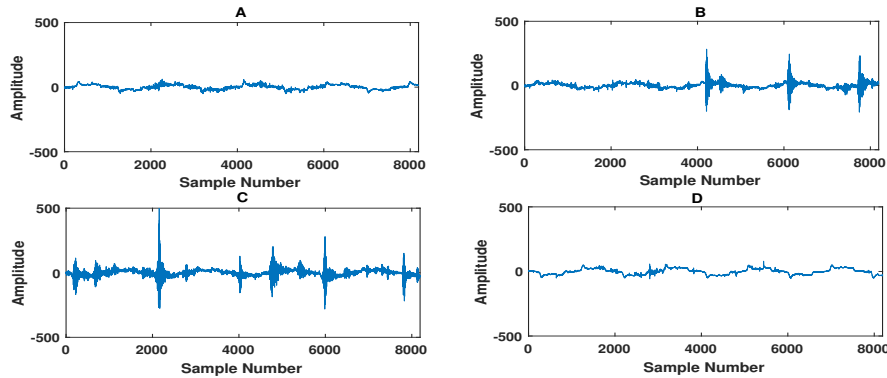


Figure 4: Examples of signals corresponding to four classes in dataset-1 A) Good condition, B) IRF, C) ORF, D) IORF

#### 4.2. Gear Fault Dataset (Dataset-2)

The dataset-2 consists of 7 classes of normal and abnormal conditions such as A: 10% fault, B: 20% fault, C: 30% fault, D: 40% fault, E: 80% fault, F: 100% fault, G: expt no load. The data is sampled at a frequency of 8.2 kHz and sampling length of 2047. Figure 5 represents the examples of signals corresponds to these 7 classes.

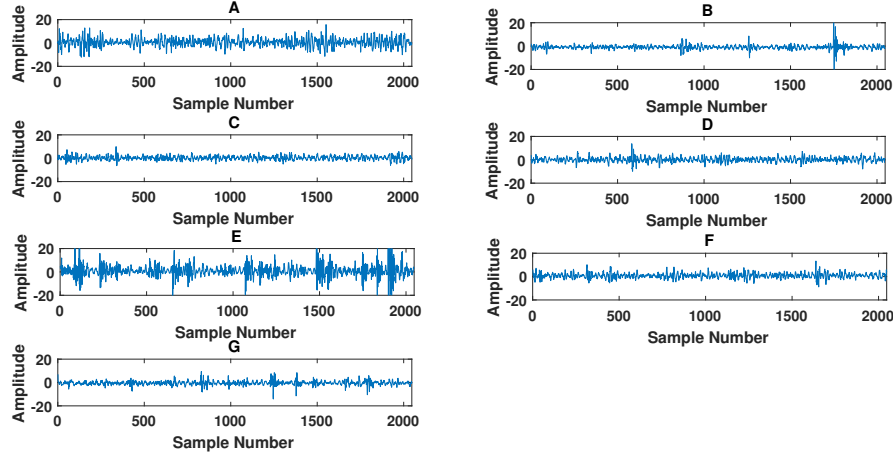


Figure 5: Examples of signals corresponding to 7 classes in dataset-2 A) 10% fault, B) 20% fault, C) 30% fault, D) 40% fault, E) 80% fault, F) 100% fault, G) expt no load

### 5. Proposed System for Machine Fault Diagnosis Monitoring

The simplified block diagram of the proposed system for machine fault diagnosis is given in figure 6. In the proposed system, the data acquisition stage contains the collection of vibration signals from machine fault simulator explained in Section 3. The vibration signals collected are belongs to two categories: 1) signals corresponds to bearing faults and 2) signals corresponds to gear faults. Further the signals are fed to the feature extraction stage where the chebfun approximation system is used for extracting the chebfun coefficients. The extracted chebfun coefficients corresponds to signals in dataset-1 is shown in figure 7. Finally this coefficients are given to the trained RLS classifier to identify the good and faulty signals.

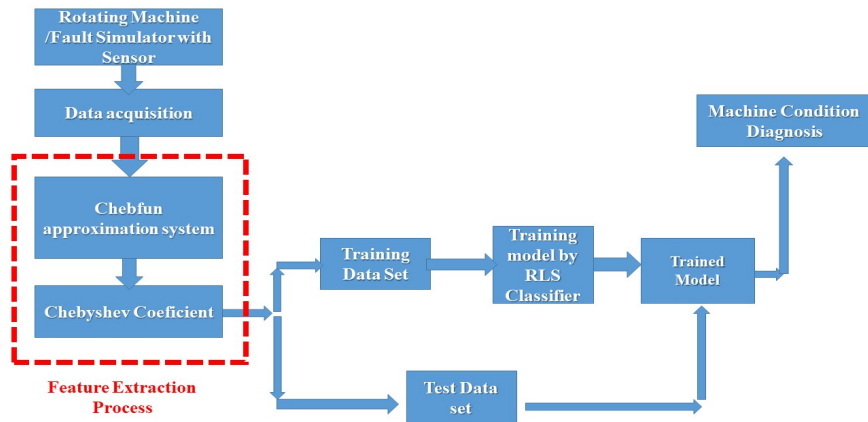


Figure 6: Block diagram of proposed machine fault diagnosis system

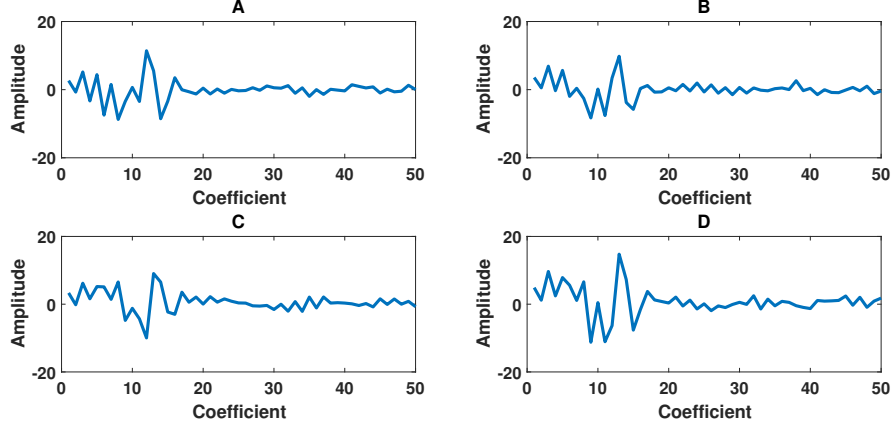


Figure 7: Examples of chebfun coefficients extracted as features using signals from dataset-1 A) Good, B) IRF, C) ORF, D) IORF

## 6. Results and Discussion

### 6.1. Evaluations using Bearing Fault Dataset (Dataset-1)

The first data set used for evaluation is bearing fault dataset (dataset-1). 100 signals from each class is generated and three splits of data are considered for evaluation. First is 20:80 data split where 20% of data is taken for testing from all the classes and remaining 80% of data is taken for training the classifier. The second data split considered is 30:70, where 30% of data is taken for testing purpose and rest 70% is used for training purpose. The third data split is 40:60, where 40% of data is taken for testing and rest for training. In the feature extraction stage, chebfun coefficients are extracted as features in 10, 20, 30, 40, 50 range and are given to RLS classifier. The performance is evaluated using two kernel functions such as linear and radial basis function (RBF). Two different cross-validation are also tried namely hold-out (HO) and leave-one-out (LOO).

Table 1: Classification accuracy for dataset-1 with different data split

Data Split	Kernel\ Coefficient	Accuracy (%)				
		10	20	30	40	50
20:80	Linear	97.5	98.54	98.03	<b>98.75</b>	95
	RBF_HO	97.5	98.12	95	97.5	98.33
	RBF_LOO	96.25	98.36	<b>98.75</b>	96.25	96
30:70	Linear	96.66	97.5	97.67	96.67	97.5
	RBF_HO	96.66	95.8	98.3	<b>98.33</b>	96.67
	RBF_LOO	96.66	97.5	<b>98.33</b>	96.67	97.5
40:60	Linear	96.87	97.5	95.62	96.25	97.5
	RBF_HO	96.87	96.87	96.85	96.25	<b>98.12</b>
	RBF_LOO	97.5	<b>98.12</b>	97.5	95	96.87

The classification accuracy for all the data split are given in Table 1. The results are noted for an average of 10 iterations. For 20:80 data split, the highest accuracy of 98.75% is obtained for Linear and RBF kernel. Also, it is observed that the accuracy is highest with 40 and 30 chebfun coefficients as features. In 30:70 data split, RBF kernel with LOO and HO cross-validation has attained the highest accuracy of 98.33% with 40 and 30 chebfun coefficients. For 40:60 splitting,

the highest accuracy of 98.12% has attained with 50 and 20 chebfun coefficients for RBF kernel. From this evaluations, it is clear that the proposed system has guaranteed an average accuracy of 98% for bearing fault discrimination.

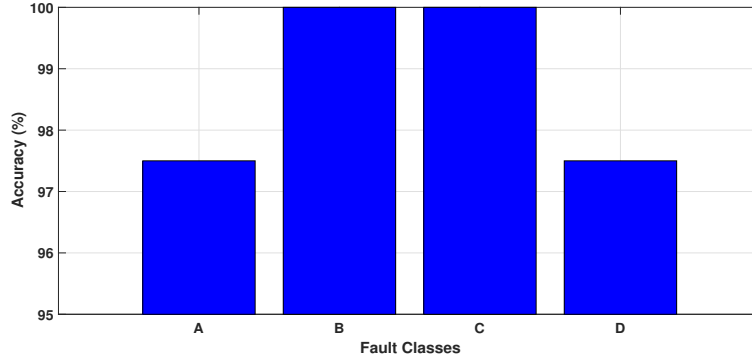


Figure 8: Illustration of class-wise accuracy for dataset-1 A) Good, B) IRF, C) ORF, D) IORF

The bar chart in figure 8 shows the class-wise accuracy obtained for dataset-1 in a single iteration with linear kernel. Here, the faulty classes (IRF, ORF) are accurately classified with 100% result. The good and IORF has attained an accuracy of 97.5%. The highest class-wise accuracy obtained confirms the superior performance of the proposed method.

### 6.2. Evaluations using Gear Fault Dataset (Dataset-2)

The second data set used for evaluation is gear fault dataset (dataset-2). It contains 7 different gear fault conditions. 60 signals from each class is generated for evaluation. The three splits of data (20:80, 30:70, 40:60) same as previously explained is considered. In this case also the chebfun coefficients are extracted as 10,20,30,40,50 range and are classified using RLS. The

Table 2: Classification accuracy for dataset-2 with different data split

Data Split	Kernel\ Coefficient	Accuracy (%)				
		10	20	30	40	50
20:80	Linear	91.66	96.42	<b>98.80</b>	97.61	95.23
	RBF_HO	91.66	92.85	90.47	97.61	91.6
	RBF_LOO	92.85	96.42	90.47	91.66	95.23
30:70	Linear	95.23	<b>96.03</b>	92.85	92.85	94.44
	RBF_HO	84.12	85.71	86.50	92.06	92.06
	RBF_LOO	88.50	84.12	92.06	87.30	86.50
40:60	Linear	89.88	91.66	82.73	<b>94.04</b>	89.88
	RBF_HO	77.97	86.90	78.57	82.73	81.54
	RBF_LOO	82.73	80.95	85.71	83.92	82.73

classification accuracy for all the data split are tabulated in Table 2. The results are noted for an average of 10 iterations. For 20:80 data split, the highest accuracy of 98.80% is obtained for linear kernel with 30 chebfun coefficient as features. In 30:70 data split, linear kernel has attained the highest accuracy of 96.03% with 20 chebfun coefficients. For 40:60 splitting, the highest accuracy of 94.04% is achieved with 40 chebfun coefficient and linear kernel. From this evaluations, it is clear that the proposed system has obtained a highest accuracy of 98% . The class-wise accuracy for a single iteration with linear kernel is shown in figure 9. From the figure

it is evident that the proposed system is able to classify the fault conditions with very high accuracy except for class B, (for class B (20% fault) accuracy is 92.75%.

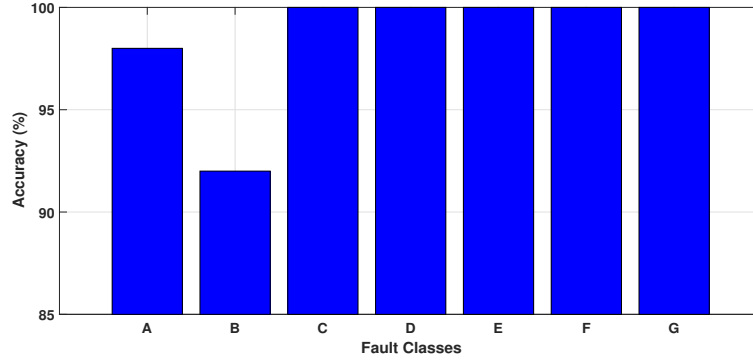


Figure 9: Illustration of class-wise accuracy for dataset-2 A) 10% fault, B) 20% fault, C) 30% fault, D) 40% fault, E) 80% fault, F) 100% fault, G) expt no load

From the experimental evaluations it is concluded that the proposed fault diagnosis system is accurate. For both bearing and gear dataset, the proposed system has achieved the best accuracy with 20:80 data split using 30-50 chebfun coefficient features.

Table 3: Performance comparison with existing approaches in terms of accuracy

Existing Approaches	Fault Category	No.of Classes	Accuracy (%)
Stat+RLS-RBF [2]	Bearing	12	96.67
Cyclo+DT [1]	Bearing	4	94
Cyclo+Non-Prob. SMO Linear Polykernel [1]	Bearing	4	97.25
Cyclo+Non-Prob. SMO Linear Polykernel [1]	Bearing	4	96.75
Cyclo+RLS-RBF [1]	Bearing	4	96.45
Stat+cSVC-Linear [5]	Bearing	12	98.03
Stat+nuSVC-Linear [5]	Bearing	12	95.58
Stat+PSVM [8]	Bearing	4	95
Stat+J48 [17]	Gear	7	85.1
Stat+NN [18]	Ball Bearing	3	95
<b>Chebfun+RLS-Linear (Proposed System)</b>	<b>Bearing</b>	<b>4</b>	<b>98.75</b>
<b>Chebfun+RLS-Linear (Proposed System)</b>	<b>Gear</b>	<b>7</b>	<b>98.80</b>

### 6.3. Comparison with Existing Methods

The classification performance of the proposed machine fault diagnosis system is compared with various existing machine condition monitoring systems and are given in Table 3. The existing systems used for comparison are stat-RLS [2], cyclo-DT [1], cyclo-SMO [1], cyclo-RLS [1], stat-cSVC [5], stat-nuSVC [5], stat-PSVM [8], stat-J48 [17] and stat-NN [18]. In stat-RLS [2] method,



fault diagnosis is done using RLS algorithm with statistical features. In [1], vibration analysis is performed using cyclostationary features with decision tree (cyclo-DT), sequential minimum optimization (cyclo-SMO), RLS (cyclo-RLS) algorithms. In stat-SVC [5] method, statistical features with c-SVC and nu-SVC classifiers is proposed for machine fault diagnosis. In stat-PSVM [8], statistical features are classified using proximal support vector machine (PSVM) for machine fault detection. In stat-J48 [17], a decision tree -J48 algorithm using statistical features with an initial feature selection procedure is adopted for classification purpose. In [18], a neural network (NN) system trained with statistical parameters are used for ball bearing classification. From Table 3, it is clear that the proposed fault monitoring system achieved a higher performance with the existing methods. This highlight that the proposed chebfun based system can be used for real time machine fault diagnosis with state-of-the-art classification accuracy.

## 7. Conclusion

In this paper, the significance of chebfun approximation system in vibration signals for machine condition monitoring is investigated. The vibration signals taken from two different machine fault setup namely bearing fault and gear fault are considered. The collected vibration signals are given to chebfun approximation system to extract chebfun coefficients as features. Further the extracted features are given to RLS classifier. The experimental evaluations are performed using various kernel functions and chebfun coefficients. The proposed system has achieved an average accuracy of 98% for both bearing and gear fault classification. Also, the performance is compared with several other existing approaches and the results are found to be satisfactory. Hence, it is concluded that the proposed machine fault diagnosis system using chebfun coefficients can be used as an effective tool for real-time machine condition monitoring.

## References

- [1] S Sachin Kumar, Neethu Mohan, Prabakaran Poornachandran, and KP Soman. Condition monitoring in roller bearings using cyclostationary features. In *Proceedings of the Third International Symposium on Women in Computing and Informatics*, pages 690–697. ACM, 2015.
- [2] Neethu Mohan, PS Ambika, S Sachin Kumar, Murugan Sai M, and KP Soman. Multicomponent fault diagnosis using statistical features and regularized least squares. *International Journal of Applied Engineering Research*, 10(20):19074–19080, 2015.
- [3] Luisa F Villa, Anfbal Reñones, Jose R Peran, and Luis J de Miguel. Statistical fault diagnosis based on vibration analysis for gear test-bench under non-stationary conditions of speed and load. *Mechanical Systems and Signal Processing*, 29:436–446, 2012.
- [4] V Sugumaran and KI Ramachandran. Fault diagnosis of roller bearing using fuzzy classifier and histogram features with focus on automatic rule learning. *Expert Systems with Applications*, 38(5):4901–4907, 2011.
- [5] V Sugumaran, GR Sabareesh, and KI Ramachandran. Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine. *Expert Systems with Applications*, 34(4):3090–3098, 2008.
- [6] M Amarnath, V Sugumaran, and Hemantha Kumar. Exploiting sound signals for fault diagnosis of bearings using decision tree. *Measurement*, 46(3):1250–1256, 2013.
- [7] Athira Chandran, T Anjali, Neethu Mohan, and KP Soman. Overlapping group sparsity induced condition monitoring in rotating machineries. In *International Conference on Soft Computing and Pattern Recognition*, pages 409–418. Springer, 2016.
- [8] V Sugumaran, V Muralidharan, and KI Ramachandran. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical systems and signal processing*, 21(2):930–942, 2007.
- [9] Zachary Battles and Lloyd N Trefethen. An extension of matlab to continuous functions and operators. *SIAM Journal on Scientific Computing*, 25(5):1743–1770, 2004.
- [10] Rodrigo B Platte and Lloyd N Trefethen. Chebfun: a new kind of numerical computing. In *Progress in industrial mathematics at ECMI 2008*, pages 69–87. Springer, 2010.
- [11] James F Whidborne. Solving optimal control problems using chebfun. In *2016 UKACC 11th International Conference on Control (CONTROL)*, pages 1–6. IEEE, 2016.

- [12] Neethu Mohan and KP Soman. Power system frequency and amplitude estimation using variational mode decomposition and chebfun approximation system. In *2018 Twenty Fourth National Conference on Communications (NCC)*, pages 1–6. IEEE, 2018.
- [13] B Ganga Gowri, KP Soman, and D Govind. Improved epoch extraction from telephonic speech using chebfun and zero frequency filtering.
- [14] Andrea Tacchetti, Pavan K Mallapragada, Matteo Santoro, and Lorenzo Rosasco. Gurls: a least squares library for supervised learning. *The Journal of Machine Learning Research*, 14(1):3201–3205, 2013.
- [15] S Sachin Kumar, B Premjith, M Anand Kumar, and KP Soman. Amrita\_cen-nlp@ sail2015: sentiment analysis in indian language using regularized least square approach with randomized feature learning. In *International Conference on Mining Intelligence and Knowledge Exploration*, pages 671–683. Springer, 2015.
- [16] S Sachin Kumar, M Anand Kumar, and KP Soman. Identifying sentiment of malayalam tweets using deep learning. In *Digital Business*, pages 391–408. Springer, 2019.
- [17] V Sugumaran, Deepak Jain, M Amarnath, and Hemantha Kumar. Fault diagnosis of helical gear box using decision tree through vibration signals. *International Journal of Performability Engineering*, 9(2), 2013.
- [18] M Subrahmanyam and C Sujatha. Using neural networks for the diagnosis of localized defects in ball bearings. *Tribology International*, 30(10):739–752, 1997.