Fault Diagnosis in Centrifugal Pump using Support Vector Machine and Artificial Neural Network

Nagendra Singh Ranawat, Pavan Kumar Kankar and Ankur Miglani

System Dynamics Lab, Indian Institute of Technology, Indore, Madhya Pradesh, India
Corresponding Author: pkankar@iiti.ac.in.

ABSTRACT

Centrifugal pumps are commonly utilized in thermo-fluidic systems in the industry. Being a rotating machinery, they are prone to vibrations and their premature failure may affect the system predictability and reliability. To avoid their premature breakdown during operation, it is necessary to diagnose the faults in a pump at their initial stage. This study presents the methodology to diagnose fault of a centrifugal pump using two distinct machine learning techniques, namely, Support vector machine (SVM) and Artificial neural network (ANN). Different statistical features are extracted in the time and the frequency domain of the vibration signal for different working conditions of the pump. Furthermore, to decrease the dimensionality of the obtained features different feature ranking (FR) methods, namely, Chi-square, ReliefF and XGBoost are employed. ANN technique is found to be more efficient in classifying faults in a centrifugal pump as compared to the SVM, and Chi-square and XGBoost ranking techniques are better than ReliefF at sorting more relevant features. The results presented in thus study demonstrate that an ANN based machine learning approach with Chi-square and XGBoost feature ranking techniques can be used effectively for the fault diagnosis of a centrifugal pump.

 Keywords: Centrifugal pump; Condition monitoring; Feature ranking; Machine learning; Artificial neural network (ANN)

INTRODUCTION

Rotating machineries such as a centrifugal pump are ubiquitous in plethora of industries ranging from oil and gas to food industry, and other industries like paper and pulp, agriculture, and wastewater treatment. Centrifugal pumps are a critical machinery in most industrial settings, where they are used for pumping and transporting different types of fluids such as water, oil and chemicals (Sakthivel et al., 2010). A pump malfunction due to the presence of fault can deteriorate system performance and affect its reliability. Therefore, it is imperative to monitor the health of the pump continuously. Impeller and bearings are the two major parts of the pump that are susceptible to failure due to erosion by slurry particles, corrosion due to reactive chemicals and improper lubrication (Kumar & Kumar, 2017). Fault detection in these components at an initial stage can help avert catastrophic failures and reduce system downtime (McKee et al., 2011). The techniques used for fault detection are categorized into model-based (Freeman et al., 2013), data-driven (Luo et al., 2015) and hybrid fault detection (Piltan & Kim, 2020). In model-based techniques, mathematical models are used to identify a fault in the system. Mathematical model of the centrifugal pump has been developed to detect the fault in the pump (Kallesoe, 2005). Data-driven technique is based on analyzing and processing the raw signals obtained from the system and has better capability to detect fault compared to a model-based approach (P. & G., 2012). In data-driven approach, the signals of different parameters are analyzed either power cells based on Cadmium Telluride (Wu et al., 2004). Till now CdTe solar cell has achieved an efficiency of only 22.1% practically (Green et al., 2020) even though the efficiency of close to 32% is possible theoretically (Agostinelli et al., 2003). This is primarily due to the absorption losses of short CdS wavelength and an interlinking individually or simultaneously; in case of interconnected parameters such as the motor's electrical signals and fluid pressure and temperature signals. In the latter case, synchronized measurements of different

parameters are performed and analyzed simultaneously to identify the fault and determine its root cause (Luo et al., 2015; Stan et al., 2018). While these signals aid in fault diagnosis, they are still an indirect signature of the actual fault condition. For a rotating machinery, the vibration signals acquired using an accelerometer reflect the true state of an actual fault, and therefore, the condition of the rotating system (Farokhzad et al., 2013). By applying intelligent machine learning (ML) -based techniques such as SVM, Decision tree (DT) and ANN (Kankar et al., 2011; Muralidharan et al., 2014; Sakthivel et al., 2010) on these vibration signals the health of a rotating system can be determined ; via feature extraction from the signal in time, frequency and time-frequency domains

(Sharma et al., 2017).To improve the computational efficiency of a ML-model, the dimensionality of the features can be decreased by selecting optimal features. Different feature ranking (FR) techniques such as Chi-square, Fisher score, Information gain, Mutual Information, ReliefF, Wilcoxon ranking, XGBoost have been employed by previous studies for selecting the optimal features (Prakash & Kankar, 2020; Sharma et al., 2017; Vakharia et al., 2016). These optimal features can be utilized to train the ML-model to detect and classify the faults in rotating machines.

In this study, the SVM and ANN models have been used to diagnose and categorize the faults in a centrifugal pump. The features embedded in the raw signals have been extracted in time and frequency domain. These feature vectors are then reduced by selecting the features using Chi-square, ReliefF, and XGBoost methods. The features selected are employed as input to these classifiers. The accuracy of two models is evaluated and compared using three different feature ranking (FR) techniques, and it is demonstrated that ANN technique is more suitable for fault diagnosis and classification.

EXPERIMENTAL SETUP

The dataset of the self-centrifugal pump used in this study is obtained from Lu et al., 2016. The accelerometer fixed on the pedestal above the motor housing has been used to acquire the vibration signal, as shown in Figure 1. The experiments were carried out at the rotational speed of 2900 rpm. The vibration signal was recorded using acceleration sensor at a sampling rate of 10239 Hz. The vibration signal was obtained under five different working conditions: normal operation, which is the base line case, and four cases with faults; impeller wear, roller wear, bearing outer race wear and bearing inner race wear. These five working conditions are listed in Table 1. For each fault condition 5 datasets were acquired, each with a sampling time of 2 s. One representative signal of the acceleration for the normal working condition and each type of faulty condition is demonstrated in Figure 2. The raw data is distributed such that there are 60 datasets for each type of faults.
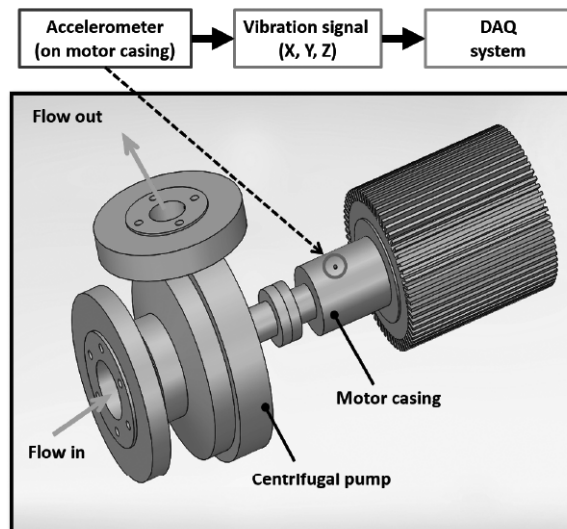
Figure 1. A schematic representation of the experimental facility used for acquiring vibration signals from the motor casing of a self-priming centrifugal pump. Location of the accelerometer for data acquisition is shown marked.
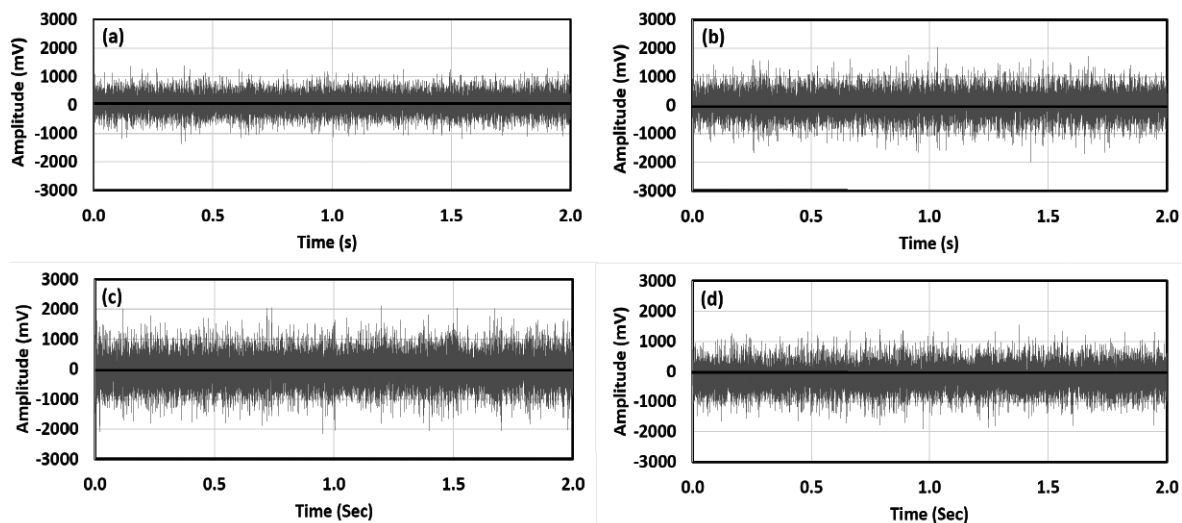
Table 1. Different condition of Centrifugal Pump

S. No	Conditions of pump
1	Normal condition
2	Bearing inner race wear
3	Bearing outer race wear
4	Bearing roller wear
5	Impeller wear

FEATURE EXTRACTION

Features are obtained from the vibration signals to detect and classify faults in the system. For monitoring system health, the features from time-domain are good indicators but to locate a particular fault in the system, frequency-domain features show better results (Sharma et al., 2017). Therefore, the accuracy of the model improves by combining the features in both these domains.

Time domain feature extraction: The different features that are obtained in the time domain of the vibration signal are as follows; Mean ($\bar{x}$): It is defined as the summation of the data divided by the number of elements in data and is given by $\bar{x} = \frac{\sum_{i=0}^{n} x_i}{n}$, Standard deviation($\sigma$): It represents the variation of the data around the average value of the data and is given by $\sigma = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}}$, RMS ($x_{RMS}$): It is depicted as the square root of the arithmetic mean of the squared data and is given by $x_{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} |x_n|^2}$, Kurtosis (K): It is defined as the degree to which the data is distributed on either of the tails in the data distribution curve i.e., the extent to which the tails of data distribution vary from tails of a normal distribution. It also shows the arrangement of dataset around its average value and is given as $K = \frac{\sum_{i=1}^{n} \frac{(x_i - \bar{x})}{n}}{\sigma^4}$, Skewness (S): It represents the symmetry of the data around its average value is given as $S = \frac{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^3}{n}}{n}$ and Crest Factor (CF): It is defined as an absolute maximum value divided by the RMS of data is given as $CF = \frac{|x_{max}|}{RMS}$.
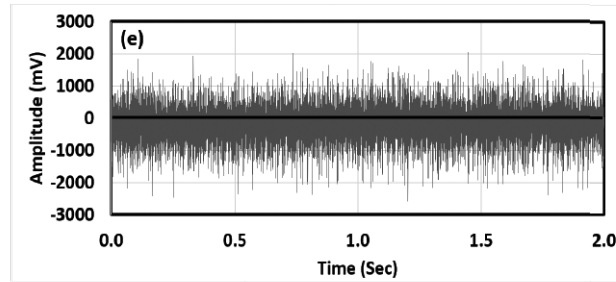
Figure. 2. Vibration signals for different working conditions of the centrifugal pump: (a) normal working operation, (b) Bearing inner race wear, (c) Bearing outer race wear, (d) Bearing roller wear, and (e) Impeller wear.

Frequency domain feature extraction: In the frequency domain, various features obtained from the signal are as follows; Peak frequency: It represent the frequency where the power in the signal is maximum, RMS frequency (f_RMS): It is depicted as the square root of the arithmetic mean of the squared value of the signal and is given as $f_{RMS} = \sqrt{\sum = \left| \frac{X(f)}{n} \right|^2}$, Spectral centroid (fc): It is depicted as the summation of frequency values which are biased by the relative spectral magnitude of every frequency component divided by the total spectral value and is given as $fc = \frac{\sum_{n}^{N=1} f(n)|X(n)|}{\sum_{n}^{N=1}|X(n)|}$, *Spectra roll-off*: It represents the frequency in the signal below which %95 of the energy is concentrated. It is the frequency at which the Power Spectral Density is more than 0.95.
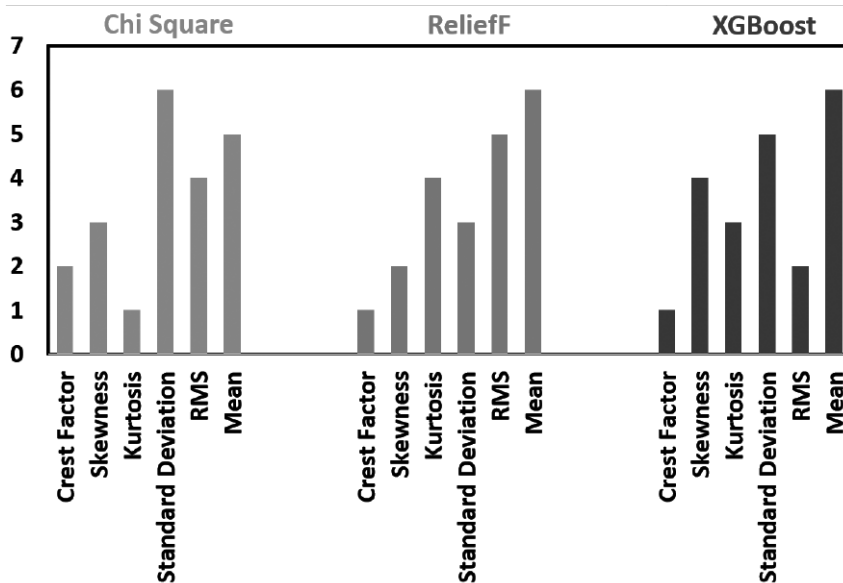

Figure 3. FR of time-domain features

FEATURE SELECTION TECHNIQUES

In each signal, the features contain information on the condition of the system, however, some features are irrelevant, and they decrease the discriminative efficiency of the model. Therefore, it is crucial to select only the optimal features that help in fault diagnosis (Prakash & Kankar, 2020). The optimal features are selected using the FR methods such as Chi-Square, ReliefF, and XGBoost. These methods are also used to decrease the dimensionality of the feature vector set for both time and frequency domain. Figures 3 and 4 show the FR results for the time-domain and the frequency-domain features respectively for three different FR techniques. After employing the ranking methods, the features selected in both the domains are combined to improve the accuracy of the model.
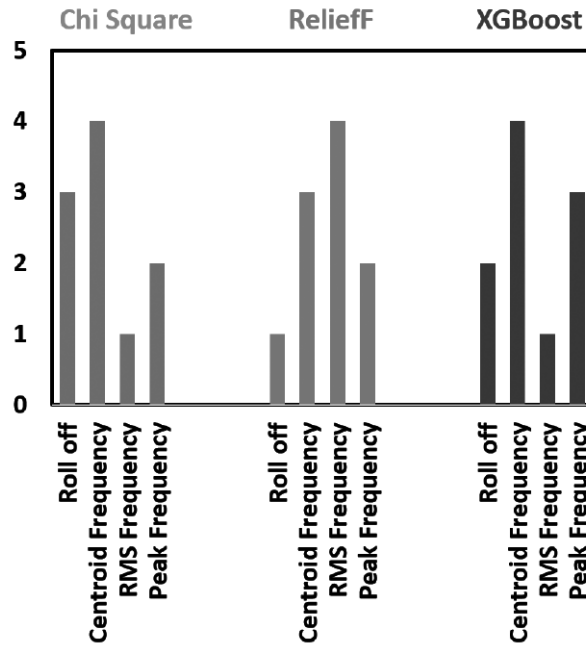
Figure 4. FR of frequency domain features

The dependency plot between the significant features selected using the FR methods and the significant feature itself is shown in Figures 5,6 and 7. The dependency plot includes a histogram when the significant feature is compared to itself. The histogram obtained is generally bimodal i.e., it has two peaks. These plots show that there is no predefined relationship between the selected features which signifies that these features are independent of each other.

The FR method decreases the dimensionality of the feature vector which reduces the computational effort required for training the classification model. The feature ranking methods used in this work are detailed below:

Chi-square: This ranking technique uses X2 statistics (Sharma et al., 2017)and the importance of the feature is evaluated independently of the concerned classes using the Chi-squared statistics. A high Chi-squared value indicates that the feature is more relevant to a specified class.

$$X2 = \sum_{i}^{A}{}_{=1} \sum_{j}^{B}{}_{=1} D_{xy} \left[Rij - \frac{\frac{Sx * Ry}{M}}{\frac{Sx * Ry}{M}} \right]^2 \qquad (1)$$

Where, M represents instances, A represents intervals and B represent classes, S_x and R_y represent the instances in x^{th} and y^{th} intervals respectively and D_{xy} represents the instances in y^{th} class and x^{th} interval.
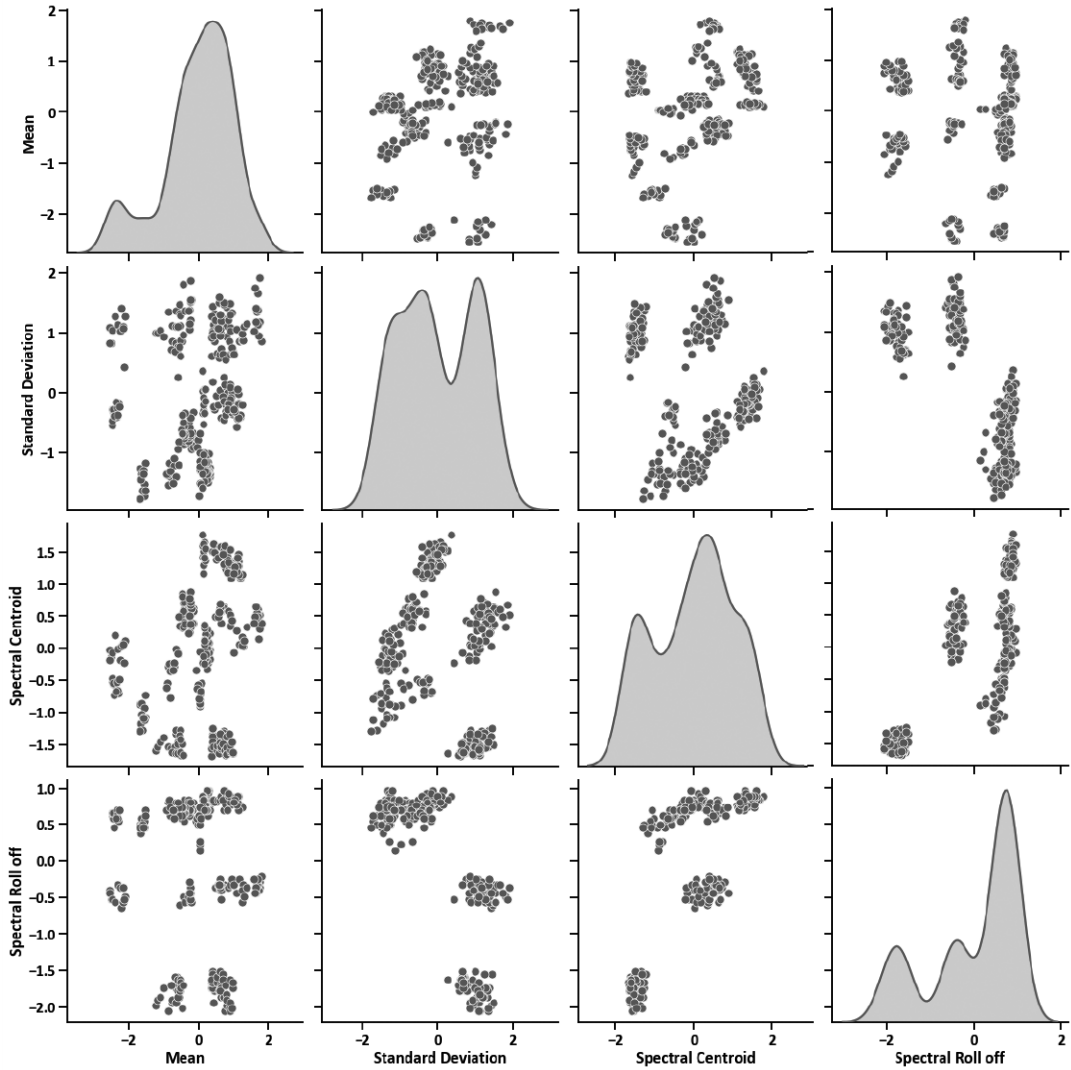
Figure 5. Dependency plot for features selected using the Chi Square method of feature extraction

ReliefF: It is a FR method that estimates the quality of a feature concerning other features. The score of features is computed in ReliefF by allocating weight to a feature depending on its capabilities to separate the sample from its nearest neighbors of the same class and the opposite class.

$$V[A] = V[A] - \Sigma^k_{j=1} \frac{\text{diff}(A,R_i,D_j)}{(m.k)} = \Sigma c \neq \text{class}(R_i) \frac{\left[\frac{P(C)}{1 - P(\text{class}(Ri))} \ \Sigma^k_{j=} 1 \ \text{diff} ((A,Ri,Lj(C))) \right]}{(m.k)} \quad (2)$$

For the occurrence R_i this algorithm hunts the kth closest member in the identical class and is named as the closest hit D_j and k^{th} closest member in the different class is named as the nearest miss L_j. The quality measure V [A] is assigned a zero value at an initial stage and it updates for the mth iteration as shown in Equation (2).
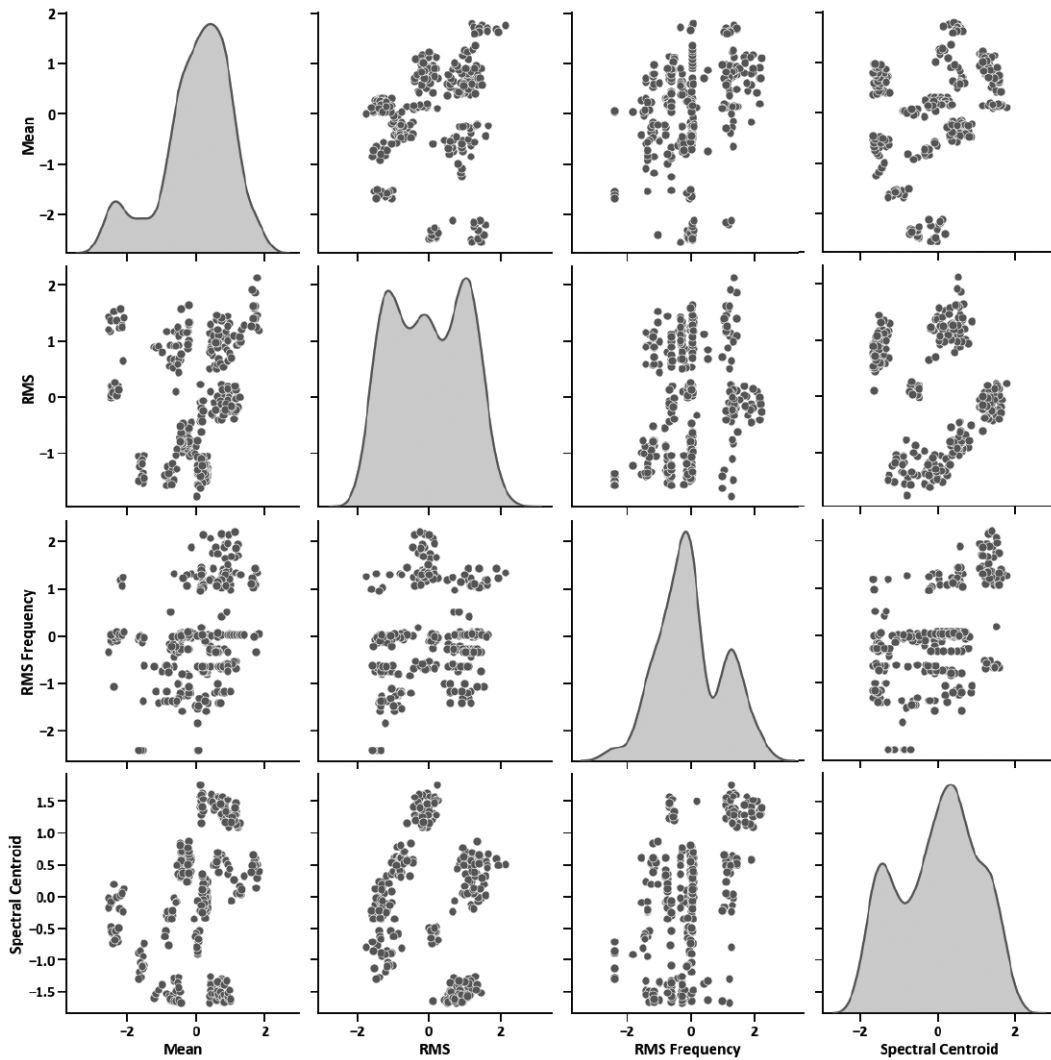
Figure 6. Dependency plot for features selected using ReliefF method of feature extraction

This algorithm rewards a feature, when it gives different values to the closest member from the other class, and it fines the feature, when it gives the same value to the closest member from the other class. Using this algorithm, the features are ranked with respect to their weights. A feature having a higher score is considered more useful as compared to others.
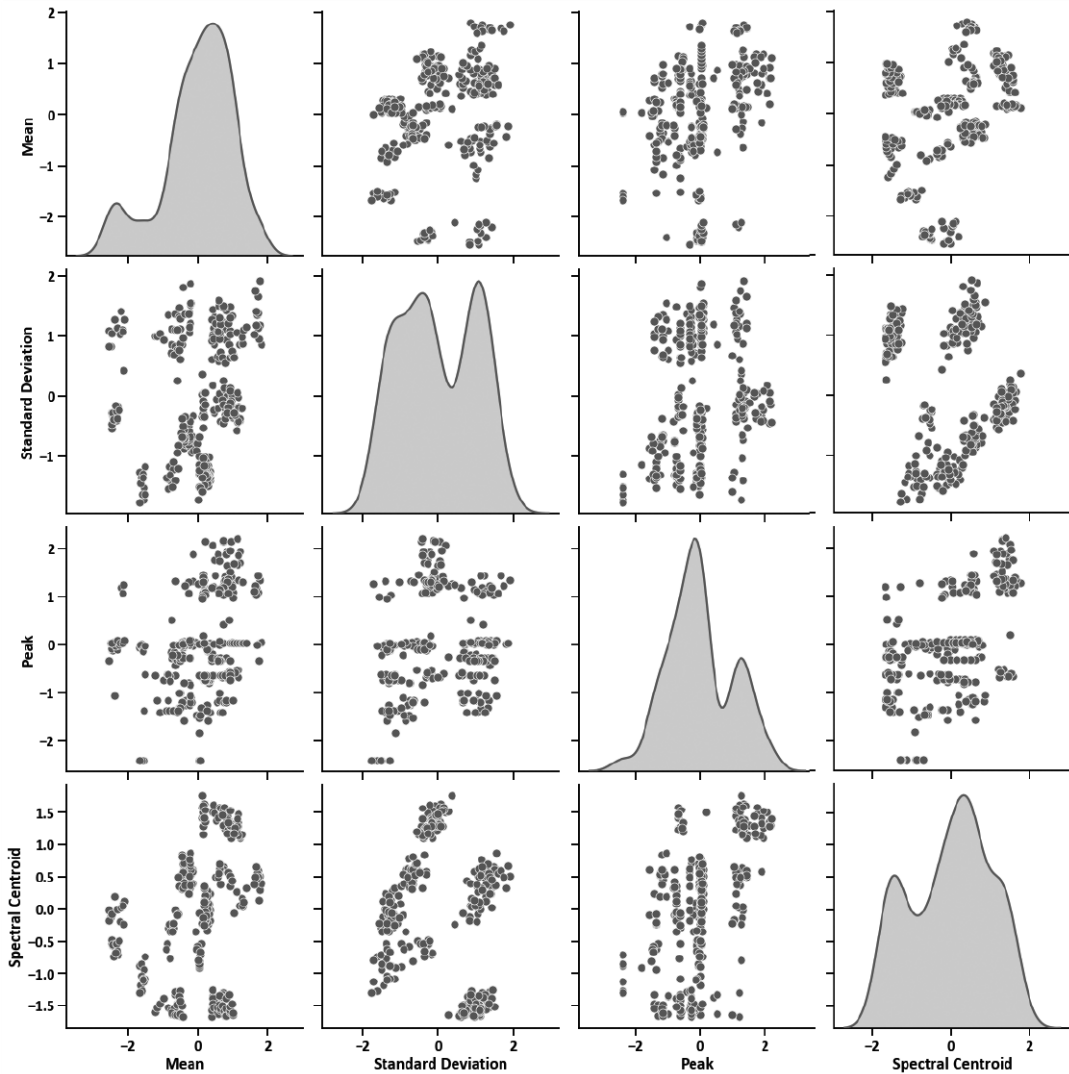
Figure 7. Dependency plot for features selected using XGBoost method of feature extraction

XGBoost: XGBoost is an ensemble technique that is based on the gradient boosting tree and efficiently constructs the boosted trees that can operate in parallel. Gradient boosting is used for FR because once the boosted decision trees are constructed then it becomes easy to retrieve information related to the feature from the model. The importance of the feature signifies the role of the feature in constructing the decision tree (Prakash & Kankar, 2020).

MACHINE LEARNING (ML) TECHNIQUES

While different ML techniques like SVM, ANN, genetic algorithm and fuzzy logic are used in engineering applications SVM and ANN (Sharma et al., 2016) are the ones that are used most widely. These two techniques are employed in this study and are briefly described below:

Support Vector Machine (SVM) – It is a supervised learning method that works on the principle of minimization of structural risk. In general, SVM is employed to solve problems related to regression and classification (Kankar et al., 2011); it can solve both linear and nonlinear classification problems. In case of a classification problem, a boundary is generated among two classes in SVM, as shown in Figure 8 (a).
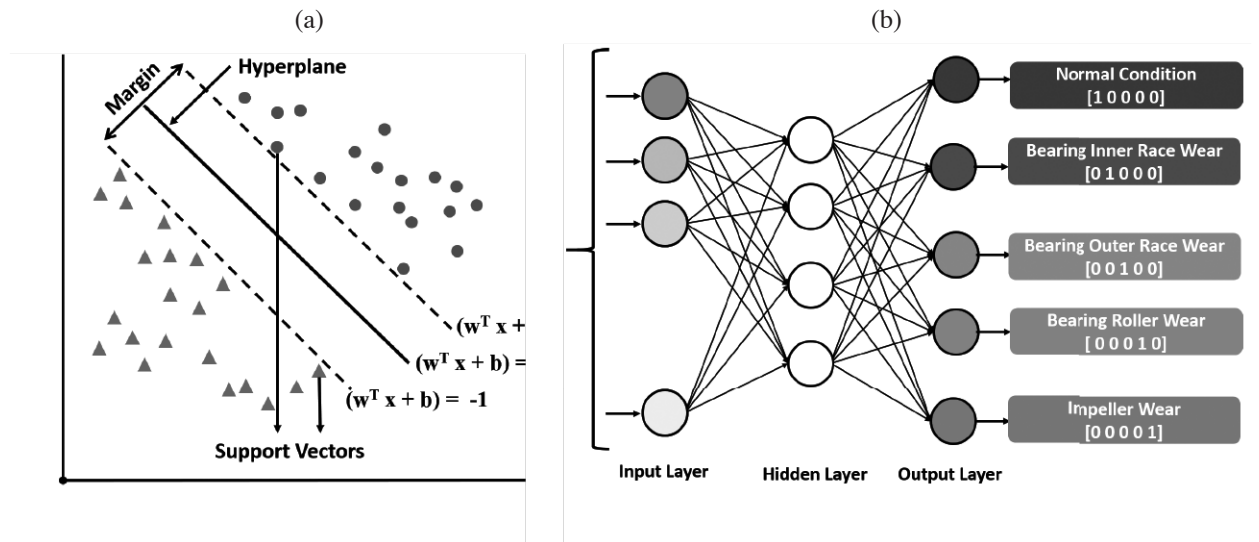
Figure 8. A schematic representation of the (a) SVM model (b) ANN Model

The bordering points which are utilized to determine the margin are named as support vectors. The key objective of this method is to increase the margin between the two points on any side of the hyperplane. This is obtained as a solution to this optimization problem as shown below:

Minimize
$$\frac{1}{2}\|w\|^2 + A\sum_{i}^{n} = 1^{\varepsilon_i} \tag{3}$$

Subject to
$$y_i\,(w'x_i + b) \geq 1\text{-}\varepsilon_i \tag{4}$$

$$\varepsilon_i \geq 0 \text{ where } i = 1,2\ldots\ldots n$$

where A represents the penalty parameter, n represents the number of samples, ε_i is slack variable, y_i is a data set, and b is the bias term.

Artificial Neural Network (ANN) – ANN is a collection of artificial nodes known as neurons. These neurons are interconnected with each other. For information processing, these neurons utilize computational or mathematical models and obtain a weighted set of inputs. Hidden neurons are used to send information from the input neurons to the output neurons. The schematic representation of ANN is shown in Figure 8 (b). ANN is an adaptive system which modifies the structure of its neurons as per the data that runs across the network. A multilayer feed-forward back-propagation algorithm is generally employed among the available architecture of the ANN for rotating machinery components such as bearings and impeller. ANN is used to solve complex problems such as fault classification and detection, which are otherwise difficult to solve using a standard computational or statistical approach. The problem solving is achieved through pattern recognition, that helps in extracting information from a complex looking data set. An artificial neuron is made of an activation function, summing function and synapses. Mathematical representation of a neuron is shown in Equation 5 below, where w_i is the interconnection weight of the input vector and the indices i and j represent the number of elements.

$$K = Z\left(\sum_{i}^{j} = 1^{w_i\,x_i} + q\right) \tag{5}$$

RESULTS AND DISCUSSIONS

In this study, different faults in centrifugal pumps such as an impeller wear, roller wear, outer race wear and inner race wear are considered. Different statistical features of the signal are obtained from both time and frequency domain. To remove irrelevant features and improve the model accuracy, three different ranking methods, namely, Chi-square test, ReliefF and XGBoost are applied to time and frequency domain separately.

Chi-square test and XGBoost ranking methods show that the mean and the standard deviation are more important features in the time domain, while the ReliefF method shows that the mean and RMS are more important as compared to other features. In the frequency domain, both Chi-square and ReliefF show agreement in terms of the dominance of spectral centroid and spectral roll off, while XGBoost displays spectral centroid and the peak frequency to be important features.
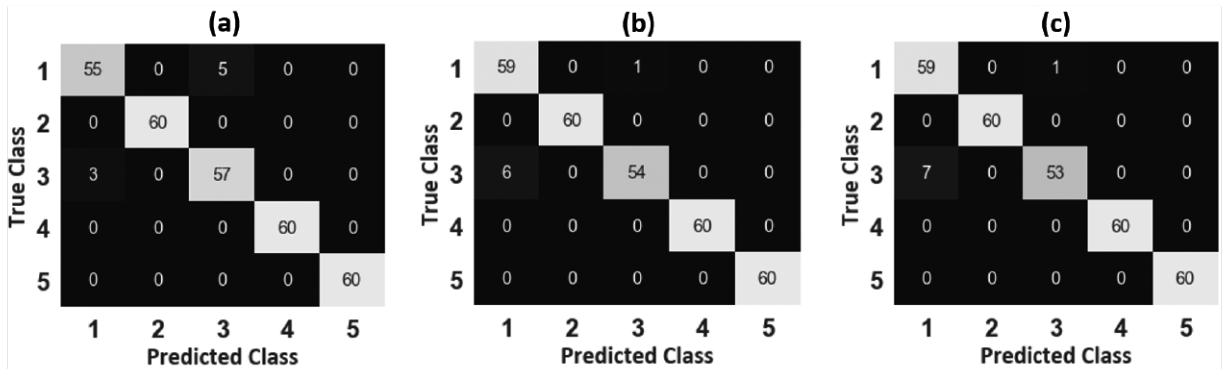
Figure 9. Confusion matrix for the SVM using different FR methods: (a) Chi square test, (b) ReliefF, and (c) XG Boost

Among the extracted features, the two best features having the highest feature importance are selected from both the domains and then combined to form four features. For analysis in the classifier, these features are utilized as inputs for training purposes. Subsequently, a comparative study of the SVM and ANN utilizing the FR method is conducted to categorize the faults in the pump. Finally, a ten-fold cross-validation is implemented to eliminate the chances of overfitting of the model. The confusion matrix for SVM and ANN with three different FR techniques is shown in Figures 9 and 10 respectively.
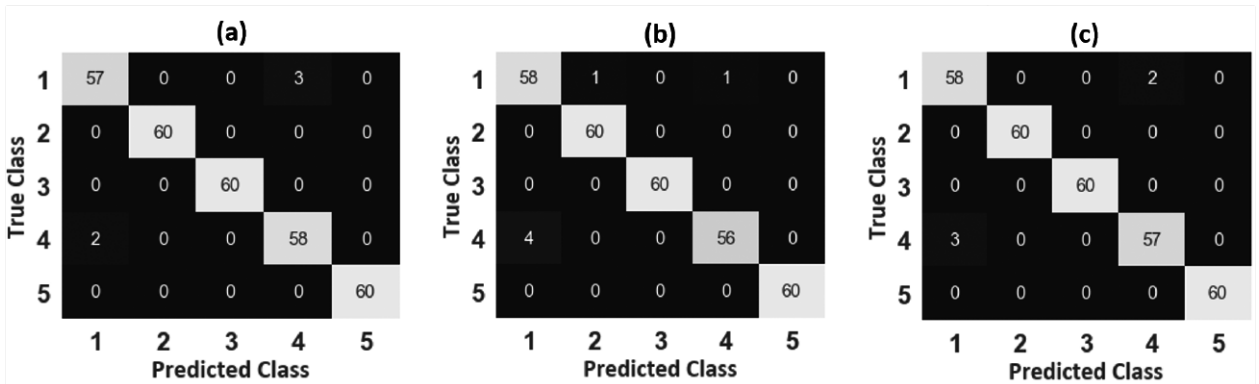
Figure 10. Confusion matrix for ANN technique using different FR methods: (a) Chi square test, (b) ReliefF, and (c) XG Boost

Table 2 shows that using the Chi-square test, the SVM and ANN reached an accuracy of 97.3 % and 98.3 % respectively. The feature selected using the ReliefF method classifies 293 instances correctly using the SVM classifier and 296 instances using the ANN.

Table 2. Accuracy of SVM and ANN techniques using different FR methods

FR method	SVM	ANN
Chi square test	97.3	98.3
ReliefF	97.7	98
XGBoost	97.3	98.3

The feature selected using the XGBoost method classifies 293 instances correctly using SVM classifier and 295 instances using the ANN. Features selection using ReliefF has a highest accuracy of 97.7 % for SVM classifier, while in case of ANN, Chi-square and XGBoost show the highest accuracy of 98.3 %. Figure 11 summarizes the entire methodology adopted for classifying faults in a centrifugal pump.
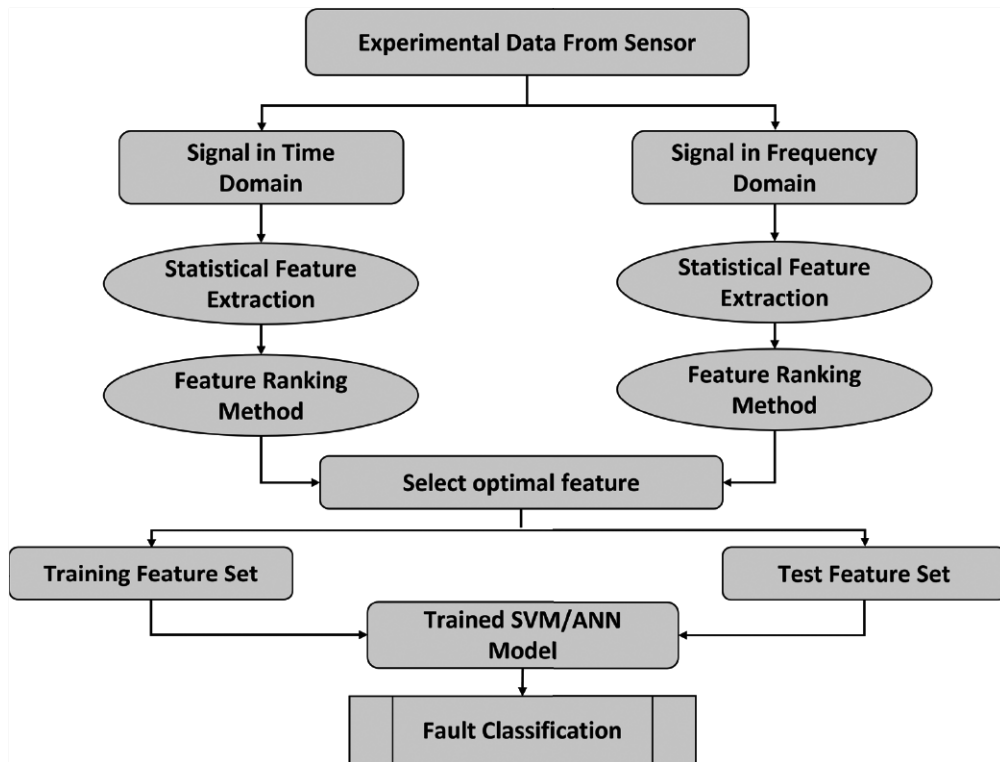

Figure 11. Methodology to classify faults in Centrifugal pump

CONCLUSION

In this study, the faults in the centrifugal pump have been classified using two ML techniques, SVM and ANN. Different sets of statistical features have been obtained from the raw vibration signal in the time and frequency domain. Three distinct FR methods have been employed to rank these features based on their importance. The extracted features are subsequently utilized to train the ML models and the performance of these models has been assessed. Based on the study, the following conclusions are drawn

a) The diagnosis and classification of faults in a centrifugal pump can be performed by analyzing its vibration signals using SVM and ANN ML techniques.

b) Both the ML techniques show good performance for fault diagnosis and classification but ANN with Chi-square test and XGBoost ranking method is found to show best results.

c) By combining the features from time and frequency domains improves the accuracy of the ML-models.

d) Feature ranking technique is an effective tool for reducing irrelevant features without compromising with the prediction accuracy of the models or increasing computational effort.

REFERENCES

Farokhzad, S., Bakhtyari, N., & Ahmadi, H. 2013. Vibration Signals Analysis and Condition Monitoring of Centrifugal Pump. Technical Journal of Engineering and Applied Sciences, 1081–1085.

Freeman, P., Pandita, R., Srivastava, N., & Balas, G. J. 2013. Model-based and data-driven fault detection performance for a small UAV. IEEE/ASME Transactions on Mechatronics, 18(4), 1300–1309.

Kallesoe, C. S. 2005. Fault Detection and Isolation in Centrifugal Pumps (Issue October).

Kankar, P. K., Sharma, S. C., & Harsha, S. P. 2011. Fault diagnosis of ball bearings using machine learning methods. Expert Systems with Applications, 38(3), 1876–1886.

Kumar, A., & Kumar, R. 2017. Time-frequency analysis and support vector machine in automatic detection of defect from vibration signal of centrifugal pump. Measurement: Journal of the International Measurement Confederation, 108, 119–133.

Lu, C., Wang, Y., Ragulskis, M., & Cheng, Y. 2016. Fault diagnosis for rotating machinery: A method based on image processing. PLoS ONE, 11(10), 1–22.

Luo, Y., Zhixiang, X., Sun, H., Yuan, S., & Yuan, J. 2015. Research on the induction motor Farokhzad, S., Bakhtyari, N., & Ahmadi, H. 2013. Vibration Signals Analysis and Condition Monitoring of Centrifugal Pump. Technical Journal of Engineering and Applied Sciences, 1081–1085.

Freeman, P., Pandita, R., Srivastava, N., & Balas, G. J. 2013. Model-based and data-driven fault detection performance for a small UAV. IEEE/ASME Transactions on Mechatronics, 18(4), 1300–1309.

Kallesoe, C. S. 2005. Fault Detection and Isolation in Centrifugal Pumps (Issue October).

Kankar, P. K., Sharma, S. C., & Harsha, S. P. 2011. Fault diagnosis of ball bearings using machine learning methods. Expert Systems with Applications, 38(3), 1876–1886.

Kumar, A., & Kumar, R. 2017. Time-frequency analysis and support vector machine in automatic detection of defect from vibration signal of centrifugal pump. Measurement: Journal of the International Measurement Confederation, 108, 119–133.

Lu, C., Wang, Y., Ragulskis, M., & Cheng, Y. 2016. Fault diagnosis for rotating machinery: A method based on image processing. PLoS ONE, 11(10), 1–22.

Luo, Y., Zhixiang, X., Sun, H., Yuan, S., & Yuan, J. 2015. Research on the induction motor current signature for centrifugal pump at cavitation condition. Advances in Mechanical Engineering, 7(11), 1–13.

McKee et al., K. 2011. A review of major centrifugal pump failure modes with application to the water supply and sewerage industries. ICOMS Asset Management Conference , Gold Coast, QLD, Australia, 1–12.

Muralidharan, V., Sugumaran, V., & Indira, V. 2014. Fault diagnosis of monoblock centrifugal pump using SVM. Engineering Science and Technology, an International Journal, 17(3), 152–157.

P., P., & G., A. 2012. Fault Diagnosis of Centrifugal Pumps Using Motor Electrical Signals. Centrifugal Pumps, June 2014.

Piltan, F., & Kim, J. M. 2020. Hybrid fault diagnosis of bearings: Adaptive fuzzy orthonormal-arx robust feedback observer. Applied Sciences (Switzerland), 10(10).

Prakash, J., & Kankar, P. K. 2020. Health prediction of hydraulic cooling circuit using deep neural network with ensemble FR technique. Measurement: Journal of the International Measurement Confederation, 151, 107225.

Sakthivel, N. R., Sugumaran, V., & Babudevasenapati, S. 2010. Vibration based fault diagnosis of monoblock centrifugal pump using decision tree. Expert Systems with Applications, 37(6),

Sharma, A., Amarnath, M., & Kankar, P. K. 2016. Feature extraction and fault severity classification in ball bearings. JVC/Journal of Vibration and Control, 22(1), 176–192.

Sharma, A., Amarnath, M., & Kankar, P. K. 2017. Novel ensemble techniques for classification of rolling element bearing faults. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 39(3), 709–724.

Stan, M., Pana, I., Minescu, M., Ichim, A., & Teodoriu, C. 2018. Centrifugal pump monitoring and determination of pump characteristic curves using experimental and analytical solutions. Processes, 6(2).

Vakharia, V., Gupta, V. K., & Kankar, P. K. 2016. A comparison of FR techniques for fault diagnosis of ball bearing. Soft Computing, 20(4), 1601–1619.