47TH TURBOMACHINERY & 34TH PUMP SYMPOSIA
HOUSTON, TEXAS | SEPTEMBER 17-20, 2018
GEORGE R. BROWN CONVENTION CENTER

A ROBUST ALGORITHM TO DETECT MULTIPLE CENTRIFUGAL PUMP FAULTS WITH CORRUPTED VIBRATION AND CURRENT SIGNATURES USING CONTINUOUS WAVELET TRANSFORM

Janani Shruti Rapur
Research Scholar
Indian Institute of Technology Guwahati
Guwahati, Assam, India

Rajiv Tiwari
Professor
Indian Institute of Technology Guwahati
Guwahati, Assam, India

Shruti is a Research Scholar at Indian Institute of Technology Guwahati (IITG). She works in the field of condition monitoring of centrifugal pumps. Prior to joining IITG, she has worked as an Applied Mechanics Engineer in Structural Analysis team at Cummins Fuel Systems, India for 2 years. She has earned her Bachelor's of Technology in Mechanical Engineering from Kakatiya University, Warangal. She has been conferred with awards such as, TCS 100 Best Student award and L&T Build India Scholarship award. Her research interests are rotor dynamics, vibrations, machine design and machine learning.

Rajiv is a Professor at Department of Mechanical Engineering, Indian Institute of Technology Guwahati (IITG). He was awarded PhD in 1997 from Indian Institute of Technology (IIT) Kanpur, India. He has been deeply involved in research pertaining to rotor dynamics. His research interests include, identification of mechanical system parameters (of rotors), fault diagnosis of machine components like bearings, gears, pumps, and induction motor and application of active magnetic bearings for condition monitoring of rotating machinery. He has authored over 170 journal and conference papers. He has guided 41 M. Tech. students and 8 Ph.D. students while 8 more are pursuing their PhD currently. He has recently published a book titled- Rotor Systems: Analysis and Identification by CRC Press, Taylor and Francis Group, USA.

ABSTRACT

Centrifugal pumps (CPs) are critical components in many industries. Their continuous availability is necessary for sustained operation of a plant. They are, however, susceptible to seizures owing to reasons such as, fluid flow abnormalities and/or mechanical component failures. These flow abnormalities not only have a direct effect on the internal components of the CP but also indirectly affect the systems delivering to or receiving flow from it. Consequently, it is crucial to recognize these faults and estimate their severity, so as to (a) curb the progression of the same and (b) improve the longevity of the CP system. The present work shows the development of a robust algorithm based on support vector machines (SVM) to classify multiple CP faults, such as suction and discharge blockages (with varying severities), impeller defects, pitted cover plate faults and dry runs in time-frequency domain using continuous wavelet transform (CWT) analysis. For the sake of classification, the CP vibration data and motor line-current data are generated for each of these faults experimentally. Furthermore, in industrial setting, CP signatures are susceptible to noise corruption due to other operating equipment in the premises. Hence, to assess the versatility of the developed methodology, the generated experimental data is further corrupted with 5%, 10% and 25% additive white Gaussian noise and used to test the algorithm. The prediction performance results thus obtained are compared to show their promise.

INTRODUCTION

The understanding of dynamics of critical rotating machines is important to ensure smooth running of the industries using them. Some

of the modern techniques of dynamic analysis of rotors are given in ref. (Tiwari, 2017). One of the critical rotating machines is a centrifugal pump (CP). CPs find wide applications, ranging from industries to domestic households owing to their versatile operation, simple design and low cost. It is estimated that a conventional chemical plant uses as many CPs as its employee headcount (Hennecke, 2000). Any untimely failure of a CP may hinder the process flow of the plant. Also, it is estimated that, 20 percent of the total energy produced around the world is utilized in driving CPs (Hart, 2002). Therefore, a continuous and efficient working of CPs is necessary for a sustained plant operation, improved personnel safety and enhanced monetary returns.

The important questions, however, are – Why does a CP fail? What are the mechanisms of failure? How to prevent or curb the failure? The CP commonly fails due to *mechanical failures* or due to *fluid flow abnormalities* or due to a combination of both of these. The details of some of the common modes of CP failure are given in Figure 1. Mechanical failures include the failure of bearings, seals, CP internal surfaces, impeller damages, unbalances, misalignments, etc. On the other hand, fluid flow abnormalities include, the flow re-circulations (suction, discharge), cavitation, fluid containing entrapped gas, dry-run, etc. Nonetheless, the authors have found that the mechanical and hydraulic CP faults are not independent of each other, and therefore, one fault can lead to the occurrence or amplification of the other (Janani & Tiwari, 2017). They form a vicious fault cycle as shown in Figure 2. Hence, it is essential that the faults are recognized in their formative stages, so that remedial measures may be timely initiated.

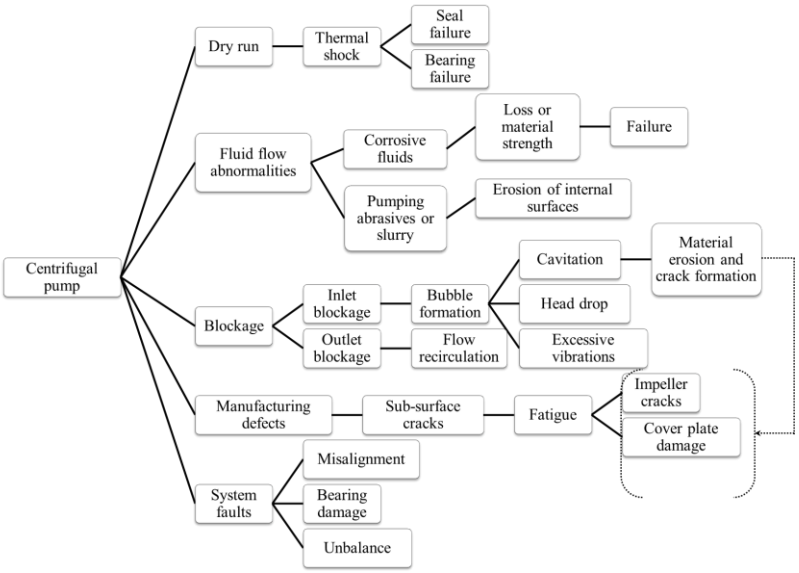

Figure 1. Different modes of CP failure

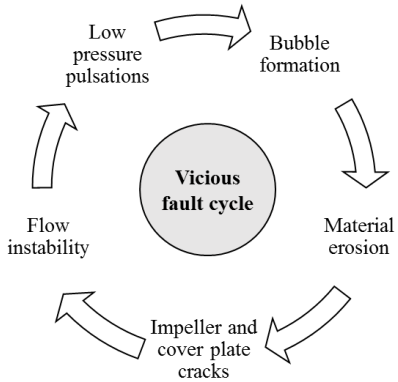

Figure 2. CP vicious fault cycle

In this study, an algorithm is developed to classify the complex CP fault combinations, as shown in Table 1. The faults considered include healthy pump (HP) condition, suction blockages (SB), discharge blockages (DB), impeller faults (IF), pitted cover-plate fault (PC), dry-run faults and their combinations as shown in Table 1.

Table 1. Faults considered in the proposed work

Fault combination	Suction Blockage (SB)	Discharge Blockage (DB)	Impeller Fault (IF)	Pitted Cover plate Fault (PC)
Healthy CP	Absent	Absent	Absent	Absent
Only SB	**Present**	Absent	Absent	Absent
Only DB	Absent	**Present**	Absent	Absent
Only IF	Absent	Absent	**Present**	Absent
Only PC	Absent	Absent	Absent	**Present**
Both IF and SB	**Present**	Absent	**Present**	Absent
Both IF and DB	Absent	**Present**	**Present**	Absent
Both PC and SB	**Present**	Absent	Absent	**Present**
Both PC and DB	Absent	**Present**	Absent	**Present**
Dry run	**Present**	Absent	Absent	Absent
Both Dry run and IF	**Present**	Absent	**Present**	Absent
Both Dry run and PC	**Present**	Absent	Absent	**Present**

A suction blockage (SB) may occur when there is an unwanted obstruction in the CP inlet pipe cross sectional area, because of, (a) clogging of the inlet pipe due to fluid impurities or (b) damages of pipe surfaces. Figure 3 shows an obstruction caused by sea weeds in the strainer of a vertical pump in a sea-water lift application. The obstruction to suction flow results in a cavitation like damage in the CP due to the pressure drop at the inlet of CP.

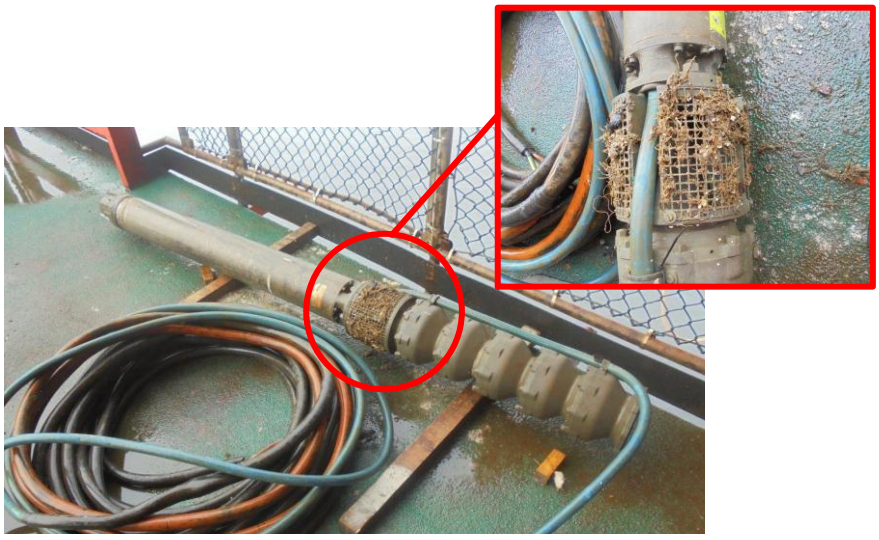

Figure 3. A sea-water lift pump with submerged motor, delivering approximately 140 m^3 sea water through a head of 14 bar suffering a strainer blockage fault

The pressure side (outlet) of the CP is generally throttled to get the required flow rate. This throttling can be considered as a discharge blockage (DB). It refers to running the CP at part load. With the throttling of the outlet, there is a decrease in flow rate and an increase of load on the impeller vanes. This may also result in flow reversal (recirculation), which results in pressure pulsations and cavitation like damage.

Most CPs are designed to operate when primed. In some cases, however, the operators do not realize the drying up of the sump or absence of priming fluid. Such cases are called as *dry run* conditions. If a CP is dry run, it results in excessive heat generation resulting in bearing and seal failures, and high vibrations.

The pitted cover (PC) and impeller faults (IFs) can be caused due to manufacturing errors and/or cavitation like damages. Whenever the smooth passage of fluid inside the CP is hindered, it results in *pseudo-fluid flow recirculation* (Janani & Tiwari, 2017).

From the previous discussions, it is clear that the causes of SB, DB, IF, PC and dry run faults are independent. However, a persistent existence of one fault can introduce or accelerate the growth of the other. Moreover, owing to the dissimilar causes of each of them they can even occur concurrently and can cause CPs catastrophic failure. Hence, fault combinations are chosen such that the acutest

conditions may be depicted. In addition, it even helps in understanding the robustness of the developed methodology to identify co-existing as well as independent CP faults.

The traditional plant maintenance techniques rely on strategies, like periodic plant maintenance or failure maintenance. The major disadvantages of these practices are,

(i) inability in identifying potential failure modes,
(ii) delayed fault detection leading to fault propagation and system failure, and
(iii) halt in the process flow of plant while maintaining it.

Therefore, the modern maintenance strategies rely on the intelligent maintenance of the plants. According to this strategy, various operating parameters of critical systems of the plant are continuously acquired and fed to the designed artificial intelligence algorithm. If any deviation of the parameter from that of the standard operation is observed, the respective failure mode is identified and the system operator is cautioned about it, so that appropriate measures can be taken to curb the fault growth. In general, vibration studies help to capture the intricate dynamics of the systems (Venkatachalam, 2014). Also, the variation on the load on the CP (due to the fault) changes the current signature of the motor. Therefore, in this paper vibration and motor line-current signatures are used to identify CP faults.

Most of the previous fault classification studies of CPs considered the mechanical and flow related faults independent of one another.

- In one of the works, the noise spectrum was used to diagnose the incipient cavitation fault in a CP (Chudina, 2003).
- Kallesoe et al. proposed a model based technique for the diagnosis of CP faults (Kallesoe *et al.*, 2006). This fault detection algorithm was based on a combination of structural analysis, observer design and analytical redundancy relation. The faults considered were clogging of CP, bearing faults, cavitation and dry run.
- Sakthivel et al. classified six faults on the CP, including the normal CP, bearing, impeller and seal faults, the combination of impeller and bearing faults, and cavitation (Sakthivel *et al.*, 2012). A fuzzy inference system based on rough set rules was developed. The performance of this classifier was compared to that of a fuzzy-ant miner and multi-layer perceptron based classifiers.
- Muralidharan and Sugumaran presented a CP fault diagnosis technique based on decision tree algorithm and continuous wavelet transform (CWT) analysis to classify faults such as cavitation, bearing defects, and a combination of impeller and bearing faults (Muralidharan & Sugumaran, 2013). In another work, the same authors team presented the CP fault diagnosis using support vector machine (SVM) algorithm (Muralidharan *et al.*, 2014).
- Sakthivel et al. demonstrated that dimensionality reduction techniques such as the PCA helps in improving the CP fault classification accuracy (Sakthivel *et al.*, 2014). They tested the method with Naïve Bayes and Bayes Net classifiers.
- Bordoloi and Tiwari presented an SVM based classification methodology to detect different stages of suction blockages in a CP (Bordoloi & Tiwari, 2017).
- Ebrahim & Javidan presented a CWT and SVM based method to classify seal and impeller faults on the CP (Ebrahimi & Javidan, 2017).
- Alabied et al. presented a classification method based on intrinsic time-scale decomposition of motor current signals for the fault diagnosis of CPs (Alabied *et al.*, 2017). The faults considered were bearing inner and outer race fault and impeller defects.

Most of the researchers considered independent existence of the mechanical and hydraulic faults. However, Perovic et al. took this problem into consideration and tried to classify cavitation faults, discharge blockage faults, impeller defects and both discharge blockage and impeller defects (Perovic *et al.*, 2001). Janani & Tiwari also considered suction blockage faults and impeller defects, together (Janani & Tiwari, 2017). But, in both of the aforementioned studies, when fault combinations were taken the classification accuracies obtained were very low. This is because it is very challenging to identify the complex fault interactions accurately. Hence, a reliable classifier needs to be designed so as to classify independent as well as coexisting faults.

There are many artificial intelligence (AI) techniques that have been previously investigated to detect CP faults, including artificial neural networks, decision trees, rough sets, deep learning, support vector machines, etc. Basic ANNs work very well when large fault database and experienced personnel are available. However, they suffer from local minima traps. Unlike the other AI techniques, the SVM works on the structural risk minimization principle and has a well-established mathematical definition (Vapnik, 1995, Widodo & Yang, 2007). Therefore, it is expected to give better learning performance as it can arrive at the near global minima. It works fine with less fault data. Due to the advantages offered by the SVM, it would be used as the AI technique to develop the fault classification methodology.

The data acquired from the CP can be used in one of the three domains, i.e. in time or frequency or time-frequency domains. The advantages or disadvantages of each of these domains rely on the behavior of the fault. In other words, it depends on the type of signal

the fault produces. For example, a bubble explosion produces a short-lived transient signal, a rotor crack or an unbalance produces a stationary-periodic signal. Hence, the selection of the domain plays an important role in the way the fault is interpreted and captured. Thus, this has a direct effect on the fault classification accuracy obtained. In the present study, a continuous wavelet transform (CWT) approach is used to pre-process the time-domain signal to accommodate both mechanical faults (which produce periodic stationary signals) and flow related faults (which produce short lived transient signals).

In this paper, the working of a developed methodology to classify distinct independent as well as co-existing mechanical and flow related faults of CPs is demonstrated. The baseline signals are corrupted with additive white Gaussian noise. Further, to check the robustness of the methodology the training and testing of the algorithm is done at various operating speeds and the performance is presented. The paper is structured as follows. Firstly, the CP lab experimentation set-up and fault types considered and combinations are explained. Next, the details of the proposed CP fault classification methodology are presented. Later on, the performance of the classification methodology is discussed, followed by the conclusions.

CENTRIFUGAL PUMP FAULT DIAGNOSIS EXPERIMENTAL SET-UP

The experimentation is performed on a machine fault simulator (MFS) setup, as shown in Figure 4. A 0.5 hp CP is mounted on the MFS and is pulley driven by a variable speed induction motor. Two tri-axial accelerometers and three motor line-current probes are used to collect the vibration and motor load variation data, respectively. One of the accelerometers is mounted on the CP casing and the second is mounted on the bearing housing location. To configure the data collection a data-acquisition system is used.

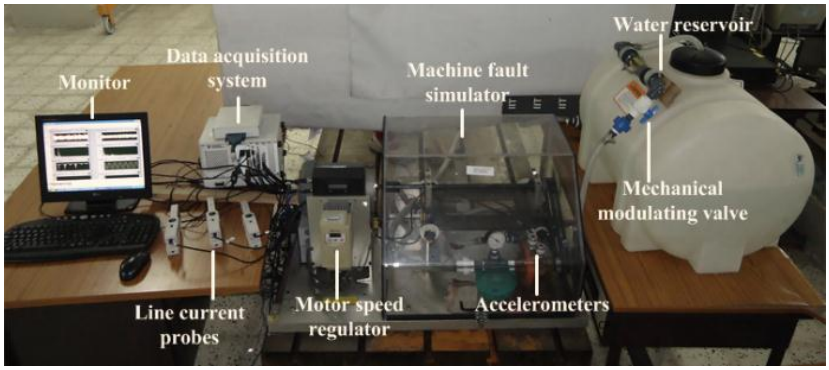

Figure 4. CP fault identification set-up. Two accelerometers are mounted one each on the CP casing and the bearing housing location. The motor line current is collected using line-current probes.

The faults are synthetically created on the CPs. Three configurations of CPs are used for this purpose. Their details are shown in Table 2. On each of these CP configurations, suction and discharge blockages are given, externally, using a mechanical modulating valve at the inlet and outlet of the CP, respectively.

Table 2. Three CP configurations

CP Type	Good impeller	Good cover plate
Healthy CP	**Present**	**Present**
CP with impeller faults	Absent	**Present**
CP with pitted cover plate faults	**Present**	Absent

These three faults are shown in Figure 5. The mechanical modulating valve can be adjusted to give the desired amount of flow restriction on the suction and discharge of the CP. The impeller is artificially given two through–through grooves on each of its vanes as impeller faults. The orientation and location of these grooves are arbitrary. The cover plate of the CP is given randomly placed pits.

A total of 33 conditions are considered on the CP. They are,

(i) healthy pump (HP) (not a fault),
(ii) suction blockage faults (SB*k*),
(iii) discharge blockage faults (DB*k*),
(iv) impeller defects (IF),
(v) impeller defects in addition to suction blockages (IFSB*k*),
(vi) impeller defects in addition to discharge blockages (IFDB*k*),
(vii) pitted cover plate faults (PC),
(viii) pitted cover plate faults in addition to suction blockages (PCSB*k*),

(ix) pitted cover plate faults in addition to discharge blockages (PCDB*k*) and
(x) dry run faults for all the CP configurations (SB5, IFSB5, PCSB5).

Here, SB*k* and DB*k* denote $(k/6)^{th}$ of suction blockage and discharge blockage, respectively, k = 1, 2, 3, 4, 5. When the CP is given $(5/6)^{th}$ flow restriction on the suction side there is almost no fluid in the CP (priming fluid), hence this case may be considered as a 'dry run' condition. A sample of the collected raw time-domain data for various faults is shown in Figure 6 and power spectrum data (frequency domain) collected for various faults is shown in Figure 7. Of these thirty-three faults investigated, the classification performance of fifteen faults is presented in this paper. They are chosen so as to approximately depict the lowest and highest severity levels of each fault. These are HP, SB2, SB5, DB1, DB5, IF, IFSB2, IFSB5, IFDB1, IFDB5, PC, PCSB2, PCSB5, PCDB1, and PCDB5.

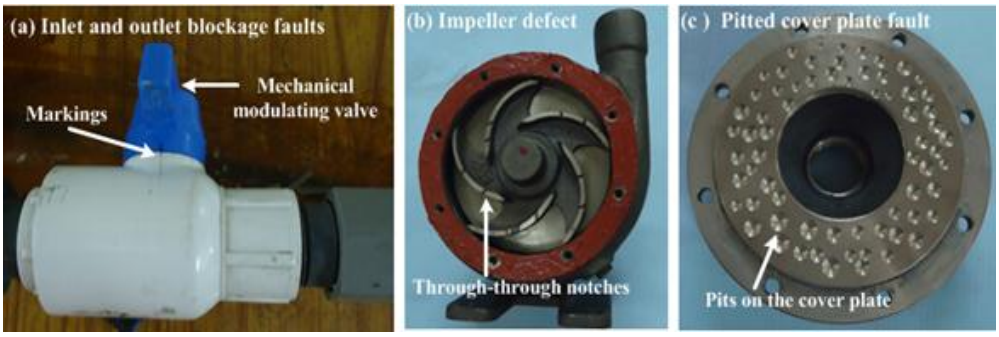

Figure 5. Faults administered on the CP, (a) mechanically modulated valve to stage different levels of suction blockages and discharge blockages, (b) impeller defects, (c) pits on the CP cover plate.

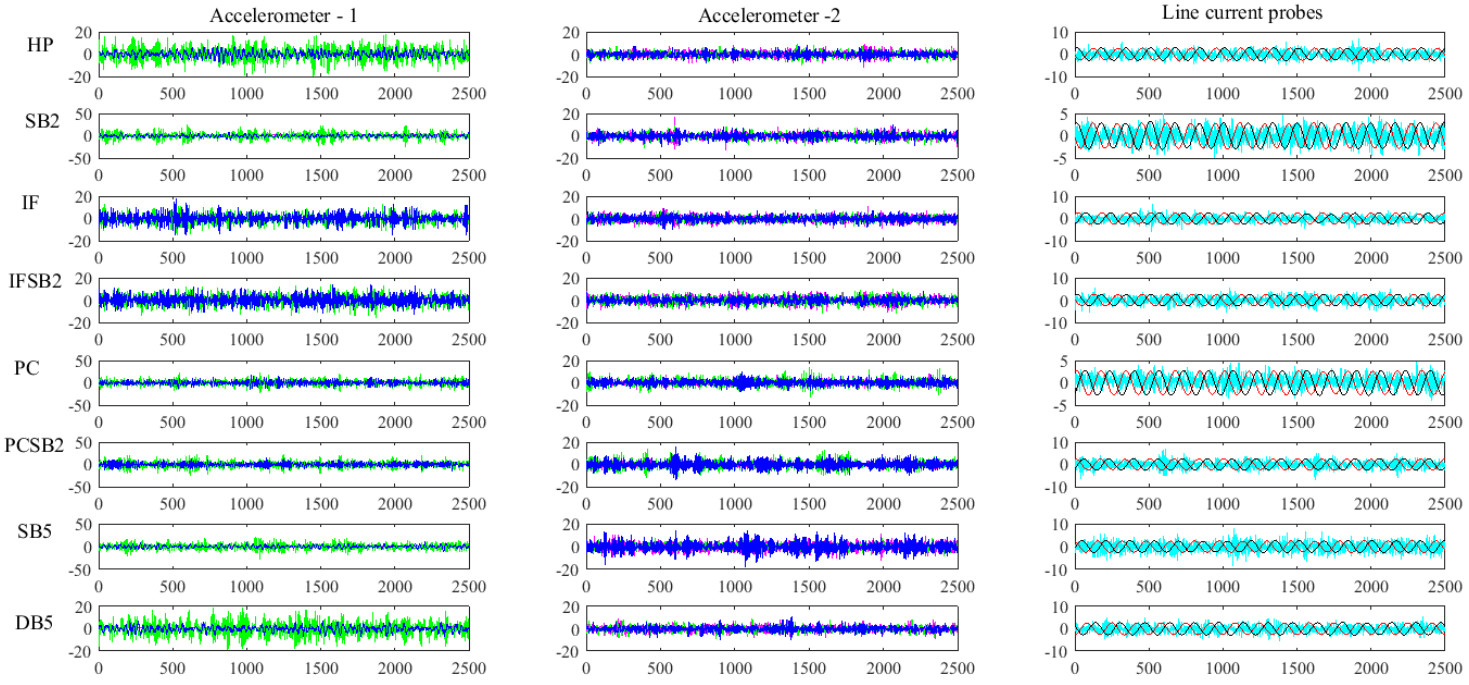

Figure 6. Raw time-domain data for various CP faults, healthy pump (HP), impeller faults (IF), combination of impeller faults and suction blockage level 2 (IFSB2), pitted cover plate faults (PC), combination of pitted cover plate faults with suction blockage level 2 (PCSB2), dry run (SB5), discharge blockage level 5 (DB5).

Some of the experimental observations made are:

- Suction blockage up to SB2 does not change the flow pattern in a CP. However, there is bubble formation observed between the SB2 and SB3 levels. The intensity of bubble formation and the size of bubbles keep increasing with the increase in the blockage severity.

- The presence of impeller defects or cover plate faults (even without any suction/ discharge blockage) causes bubble formation.
- The presence of discharge blockage reduces the pump discharge significantly and increases the vibrations.

It is witnessed experimentally that the fault manifestation (or the physical behavior of the fault) changes enormously with the changing operating speed of the CP. But, the developed method should be robust to the varying conditions of the CP. Hence, data from various operating speeds of the CP have been collected and the algorithm is tested with data from each of the speeds. The CP is therefore run from 30 to 65 Hz speed in steps of 5 Hz.

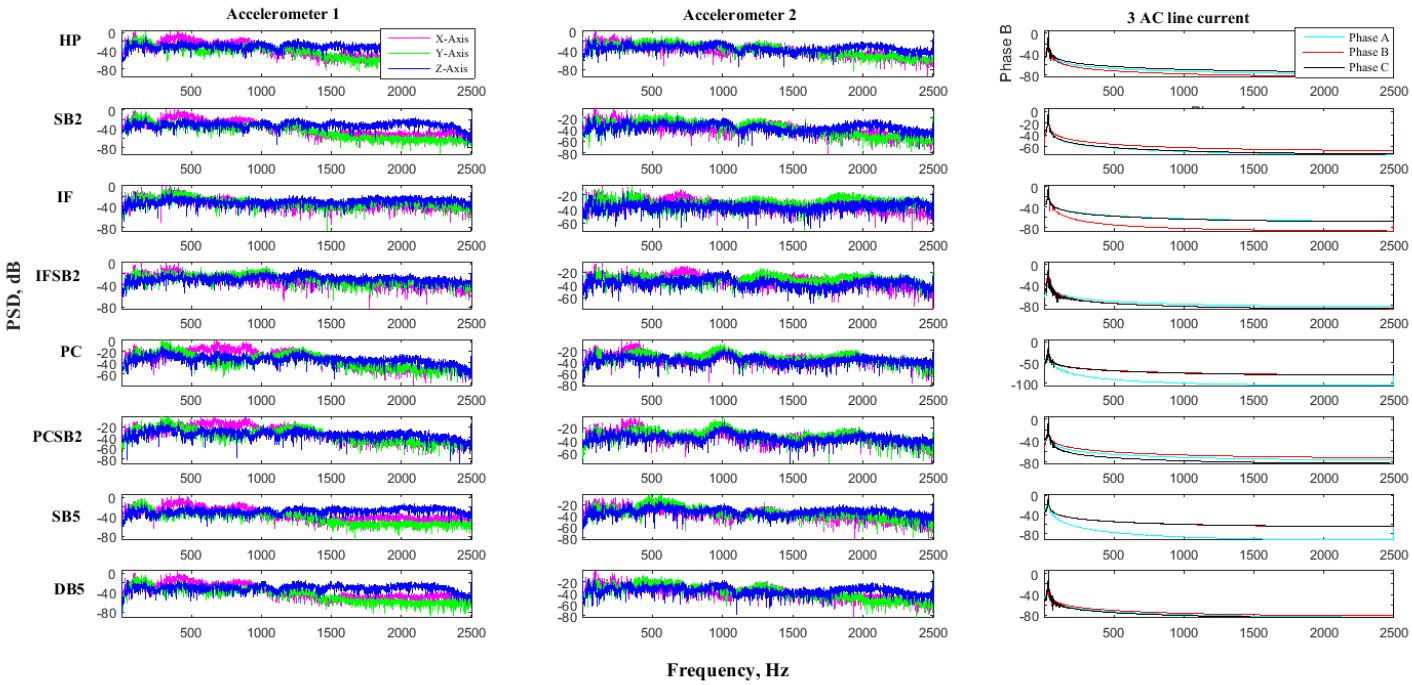

Figure 7. Raw frequency-domain data for various CP faults, healthy pump (HP), impeller faults (IF), combination of impeller faults and suction blockage level 2 (IFSB2), pitted cover plate faults (PC), combination of pitted cover plate faults with suction blockage level 2 (PCSB2), dry run (SB5), discharge blockage level 5 (DB5).

The data is sampled at two different sampling rates,

(i) 20 kHz sampling rate and 2000 sampling points per dataset which gives a Nyquist frequency of 10 kHz and data resolution of 10 Hz in frequency domain, 300 non-overlapping datasets acquired.

(ii) 5 kHz sampling rate with 5000 sampling points per dataset which gives a Nyquist frequency of 2.5 kHz and data resolution of 1 Hz in frequency domain, 150 non-overlapping datasets acquired.

These two sampling rates are chosen so as to demonstrate the role of data sampling resolution on the ease of identifying the CP system faults. Figure 8 shows the feature mapping of different faults derived from time-domain CP fault data. In the figure the x-axis is mean value and y-axis is reciprocal of standard deviation. It can be seen that with low-frequency resolution data sampling (Figure 8(a)) there is a lot of overlap of feature values for different fault conditions, however, when a higher-frequency resolution data sampling (Figure 8(b)) is chosen the faults form distinct clusters. This is because more intricate fault information is captured (especially the fault transients produced by the flow instabilities).

CENTRIFUGAL PUMP FAULT CLASSIFICATION METHODOLOGY

In general, the characteristic frequencies of faults that have stationary periodic fault signatures can be captured using the power spectrum or the fast Fourier transform (FFT) apprach. However, hydraulic faults produce transient signals. FFT considers the signal to be stationary and so may not be able to capture the transients in a signal. Therefore, while dealing with a combination of mechanical and hydraulic faults, a time-frequency domain analysis may be useful.

A brief description of continuous wavelet transform (CWT), support vector machine (SVM) algorithm and the fault classification methodology is presented in this section. The detailed account on CWT can be seen in ref. (Führ, 2005, Young, 2012) and that of SVM can be seen in ref. (Vapnik, 1995, Vapnik, 1999, Widodo & Yang, 2007).

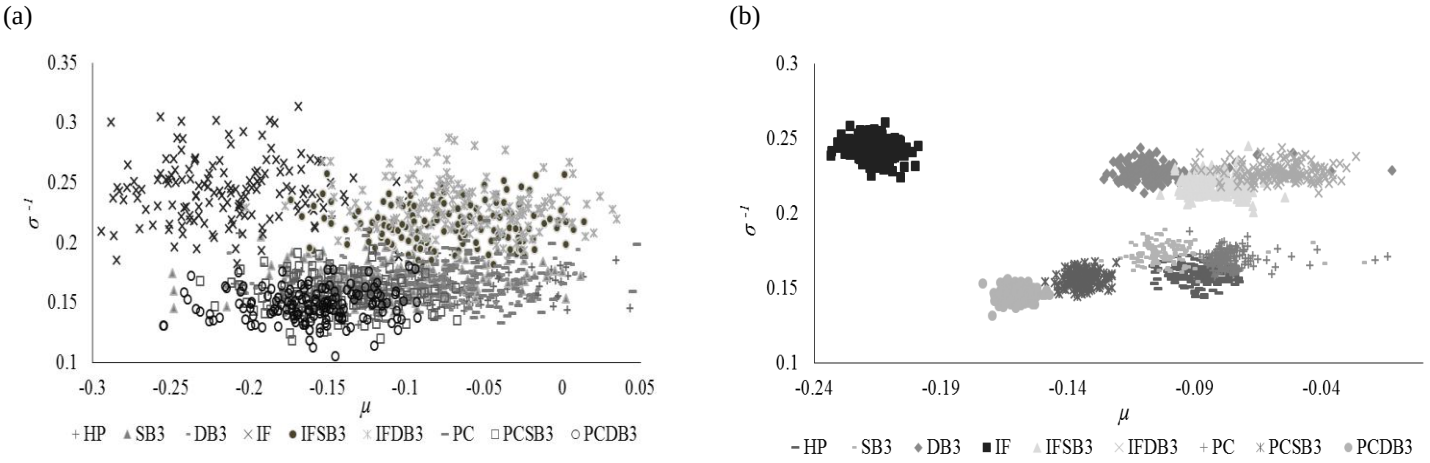

Figure 8. (a) Reciprocal of standard deviation (σ^{-1}) versus mean (μ) value for low frequency resolution data, (b) Reciprocal of standard deviation versus mean value for low frequency resolution data. Healthy pump (HP), suction blockage level 3 (SB3), discharge blockage level 3 (DB3), impeller fault (IF), combination of impeller fault and suction blockage level 3 (IFSB3), combination of impeller fault and discharge blockage level 3 (IFDB3), pitted cover fault (PC), combination of pitted cover fault and suction blockage level 3 (PCSB3), combination of pitted cover fault and discharge blockage level 3 (PCDB3)

Continuous Wavelet Transform (CWT)

Projection of a time signal onto wavelet basis functions is called '*wavelet transform*'. It is a time-frequency signal analysis method, which is widely being researched in recent years. In Fourier transform sine and cosine functions are used as the basis functions, these functions are global in time domain but are localized in frequency domain. However, in wavelet transform, 'wavelet' functions which are localized both in time and frequency domains are used. That is, wavelets are short lived functions. The choice of the mother wavelet is very important, to obtain the correct features from the signal. In the present work a wavelet from the biorthogonal family of wavelets, bior 3.3 has been chosen. The wavelet and scale functions of bior 3.3 are shown in Figures 9(left) and 9(right), respectively.

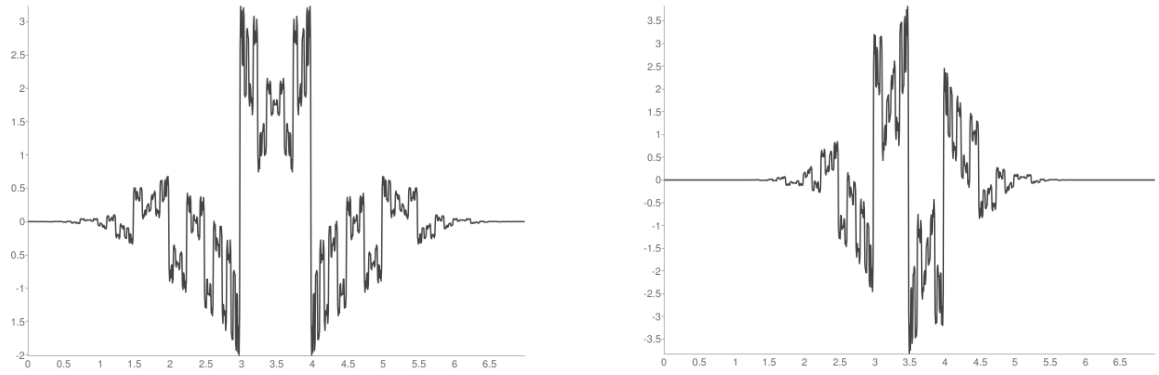

Figure 9. Amplitude versus time, (left) bior 3.3 wavelet function, (right) bior 3.3 scaling function

Support Vector Machine (SVM) Classifier

The SVM is essentially a non-probabilistic, binary classifier. In a binary data classification, as the name suggests there are two classes (types) of data, as shown in Figure 10(a). The circles represent data of class-1 and the triangles represent data of class-2. The objective of a classifier is to segregate the two classes of data and SVM uses a line to separate the two types of data. There could be infinite

orientations of a line that can separate these two classes of data (shown by dashed lines in Figure 10 (a)). But, the SVM constructs a *hyper plane* that maximizes the margin between the two classes of data, thereby dividing them into positive class (class-1) and negative class (class-2), as shown in Figure 10(b). The data points nearest to the plane, which divides the two classes of data are coloured black in Figure 10 (b). These are called *support vectors*, represented as $\mathbf{x}_n$. The margin may be given as $\|\mathbf{w}\|^{-2}$, where $\mathbf{w}$ is a vector normal to the hyper plane. Hence, in the linearly separable case, the given data input vector $\mathbf{x}_i$ ($i = 1, 2 \ldots M$), where, M is the number of samples, for finding the optimal hyper plane, can be solved using the following constraint equations,

$$\min\left[\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{M}\xi_i\right] \tag{1}$$

$$\text{Subject to,} \quad y_i(\mathbf{w}\bullet\mathbf{x}_i - b) = 1 - \xi_i \tag{2}$$

where, $i = 1, 2, \ldots, M$ and y_i takes the values +1 or -1 for the positive and negative classes, respectively, C is the error penalty and ξ_i accounts to the noise in slack variables.

Unfortunately, a linear classification of data is not always plausible. To handle such cases, the SVM transforms the data vectors into a complex dimensional feature space where the data can be linearly classified. This is done with the help of *kernel* functions. Kernel function returns the inner product of the feature space mappings of the original data points (Widodo & Yang, 2007). There are numerous predefined kernel functions available, viz. the linear, polynomial, Gaussian RBF, etc. The choice of a kernel functions and their parameters, plays a deciding role in the accuracy of classification obtained. Based on the literature (Azadeh *et al.*, 2013), in the present work a Gaussian RBF kernel is used.

In order to select optimum kernel parameters a *h*-fold cross-validation (CV) technique is adapted. According to this approach, the training data of the algorithm is distributed into '*h*' sub sets. Among these *h* subsets, (*h*-1) subsets are used for training the algorithm and one is used for testing. This process is repeated for all the *h* sub-sets. From the repeated training and testing, those kernel and SVM parameters, which give the best classification accuracies are chosen. These are then used to train the final algorithm. This method helps better generalization of the algorithm. In the present work a 5-fold CV technique is employed.

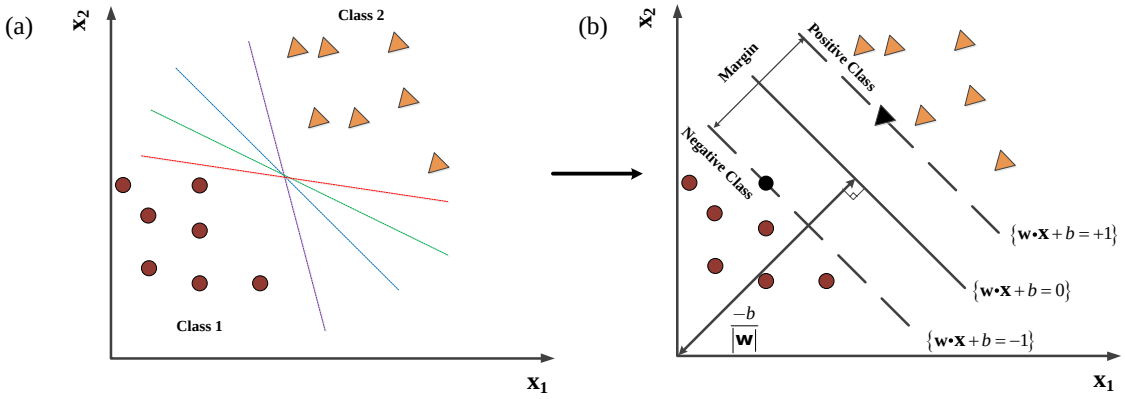

Figure 10. (a) Circles represent class-1 data and triangles represent class-2 data, dotted lines represent multiple orientations of lines that can separate the two classes of data, (b) SVM classifier with optimal orientation of hyperplane.

As previously addressed, SVM is a binary data classifier. But what if there are more than two qualities of data? To classify multi-class data (more than two types of data, ex: the case where we are trying to identify many faults on the CP), three approaches are used, they are, one versus one (OVO), one versus all (OVA), and direct acyclic graph (DAG) (Hsu *et al.*, 2003). In all the three cases, the multi-class classification model is converted to a number of binary classifications. In the present work OVO approach is used for the multi-class fault classification.

Fault classification methodology

The raw data that is acquired from CP cannot be directly input to the SVM classifier. Useful statistical features that capture the fault characteristics at range of operating conditions of the CP are extracted from the raw data. The statistical features used in the present work are the standard deviation (σ), reciprocal of standard deviation (σ^{-1}), and root sum of squares (RSS). These features have been selected using the wrapper model, which is an accuracy based technique to select the best features. All of the features mentioned have standard mathematical definitions, except σ^{-1}. This feature was shown to be useful in the CP fault diagnosis by the authors in ref.

(Rapur & Tiwari, 2017, Rapur & Tiwari, 2018).

To check the adaptability of the algorithm the signals generated from the experiments are corrupted using additive white Gaussian noise (AWGN). *White* refers to a uniform power signal across the wide frequency band for the information system. This noise has a Gaussian normal distribution with zero-mean. This model has been chosen to take into account noisy signals emerging in a broad frequency band. Also, in general any random process is estimated to have a Gaussian distribution of events. But, as the data is experimentally generated it has inherent noise. Nevertheless, noise has been added to: (a) understand how much more corruption of data can the algorithm handle, and (b) to simulate the real-world pump operation where the amount of noise interference would be much more than that in laboratory generated data.

A noisy signal is,

$$S = S_0 + N \tag{3}$$

where S_0 represents the original raw signal and N signifies the signal-dependent noise. The amplitude of the noise is of Gaussian distribution and depends on the desired signal-to-noise ratio (SNR). The SNR in dB is defined as,

$$\mathrm{SNR}_{dB} = 10\log_{10}\left(\frac{A_S}{A_N}\right)^2 \tag{4}$$

where A_S and A_N are, respectively, the amplitude of the signal and the noise. In the present work a noise of 5%, 10% and 25% are proposed to be added to corrupt the signal. This means that a SNR_{dB} of 26.02, 20 and 12.041 are used. The amount of noise has been restricted to 25% because it is expected that the sensors will be positioned close to the CP and therefore further corruption of data is avoided.

The raw data generated from the CP is collected, then, CWT coefficients are extracted from it up to scale 64. Then the aforementioned statistical features are extracted from these coefficients. Then, 50% of the data is randomly divided into training and the rest 50% into testing data. Using a 5-fold CV technique the best kernel and SVM parameters are selected. Once they are selected the classifier is ready. The test data is used to validate the algorithm. In case the data is corrupted, then the aforementioned algorithm may not be effective. The CWT coefficients and the features are directly extracted from the training data. The testing data is corrupted using the prescribed amount of noise (5%, 10% or 25% AWGN). From this corrupted data, CWT coefficients and statistical features are selected. A range of kernel and SVM parameters are given to the classification algorithm, the best combination is then selected parametrically.

By now, it is clear that there needs to be a number that quantifies the extent of fault classification. To compute that a term called - classification accuracy (C_a) is defined. It is given as,

$$C_a = \frac{\text{Number of correctely classified data points}}{\text{Total number of data points}} \tag{5}$$

Apart from C_a another term called overall classification accuracy (O_a) is defined. This is the average classification accuracy over the entire speed range. This gives a bird's eye view about the performance of the classifier.

CENTRIFUGAL PUMP FAULT DIAGNOSIS PERFORMANCE

Initially, the classification performance of the algorithm is tested using the baseline data. Here it must be noted that 'baseline' data does not mean that the data is free from noise. It only means that the experimental data is used as it is and no noise is added to it. For this case, optimized kernels as well as SVM parameters are selected using a 5-fold cross-validation technique. The results of the classification are presented in Table 3.

From the table, it can be seen that SB5, DB1, IFSB5, IFDB1, IFDB5, PCSB5, and PCDB5 faults are giving a 100% classification at all the CP running speeds. The classification accuracy is lowest at 40 Hz CP speed, and DB5 fault shows a misclassification of 10%. DB5 fault shows the least individual overall classification. Most of times, the data is misclassified into a fault case, which has close resemblance/ behaves similar to that of its original class. For example, SB2 ($2/6^{th}$ suction blockage fault) is misclassified to the HP

condition. This may be because SB2 closely behaves as the HP condition and thus feature values for both the faults may be close. The overall classification accuracy for the baseline condition is 99.53%.

Next, the algorithm is trained with baseline data and tested with 5% AWGN (noise), 10% AWGN and 25% AWGN data independently. 5% AWGN case gives an overall classification of 100%. The results of fault classification are shown in Table 4 (10% AWGN) and Table 5 (25% AWGN). To get a birds-eye view, the results classifications are compared with that of no-noise condition and are shown in Figure 11. It is interesting to observe that with the addition of white Gaussian noise the classification accuracy has not dropped significantly. The overall classification accuracies obtained for 5% AWGN, 10% AWGN and 25% AWGN are 100%, 99.9% and 99.4%, respectively. From the results shown for the baseline and corrupt data classification, it seems as if the addition of noise has improved the classification accuracy. But in an engineer's perspective there is not much difference between 99.5% and 100% classification accuracy. The standard deviation feature plot for the baseline, 5% AWGN, 10% AWGN and 25% AWGN is plotted in Figure 12. It can be observed that the addition of noise of up to 10% does not shift the feature values much. However, with the addition of 25% AWGN noise there is a perceivable amount of shift in feature values. Nevertheless, the selected features are versatile and do not differ considerably with the addition of noise. Thus, it explains the high classification percentage.

Table 3. Results of classification for baseline data.

Train/ Test Speed (Hz)	**Fault classification accuracy**															
	HP	**SB2**	**SB5**	**DB1**	**DB5**	**IF**	**IFSB2**	**IFSB5**	**IFDB1**	**IFDB5**	**PC**	**PCSB2**	**PCSB5**	**PCDB1**	**PCDB5**	C_a
30	**98**	100	100	100	**100**	**100**	**100**	100	100	100	**100**	**98**	100	**94**	100	99.3
35	**98**	100	100	100	**100**	**100**	**100**	100	100	100	**100**	**100**	100	**98**	100	99.7
40	**100**	**94**	**100**	**100**	**90**	**100**	**100**	**100**	**100**	**100**	**100**	**98**	**100**	**98**	**100**	**98. 7**
45	**98**	100	100	100	**98**	**100**	**100**	100	100	100	**98**	**100**	100	**98**	100	99.5
50	**100**	100	100	100	**98**	**100**	**98**	100	100	100	**98**	**100**	100	**100**	100	99.6
55	**100**	100	100	100	**100**	**100**	**100**	100	100	100	**100**	**100**	100	**98**	100	99.9
60	**100**	100	100	100	**100**	**98**	**100**	100	100	100	**98**	**100**	100	**100**	100	99.7
65	**100**	100	100	100	**100**	**98**	**100**	100	100	100	**100**	**100**	100	**100**	100	99.9

Healthy pump (HP), suction blockage level 2 (SB2), dry run (SB5), discharge blockage level 1 (DB1), discharge blockage level 5 (DB5), impeller fault (IF), combination of impeller fault and suction blockage level 2 (IFSB2), combination of impeller fault and dry run (IFSB5), combination of impeller fault and discharge blockage level 1 (IFDB1), combination of impeller fault and discharge blockage level 5, pitted cover fault (PC), combination of pitted cover fault and suction blockage level 2 (PCSB2), combination of pitted cover fault and dry run (PCSB5), combination of pitted cover fault and discharge blockage level 1 (PCDB1), combination of pitted cover fault and discharge blockage level 5 (PCDB5).

Table 4. Results of classification for 10% AWGN data

Train/ Test Speed (Hz)	**Fault classification accuracy**															
	HP	**SB2**	**SB5**	**DB1**	**DB5**	**IF**	**IFSB2**	**IFSB5**	**IFDB1**	**IFDB5**	**PC**	**PCSB2**	**PCSB5**	**PCDB1**	**PCDB5**	C_a
30	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
35	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
40	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
45	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
50	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
55	**100**	**100**	**100**	**100**	**99.6**	**100**	**100**	**100**	**100**	**93.5**	**100**	**100**	**100**	**100**	**100**	**99.5**
60	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
65	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Healthy pump (HP), suction blockage level 2 (SB2), dry run (SB5), discharge blockage level 1 (DB1), discharge blockage level 5 (DB5), impeller fault (IF), combination of impeller fault and suction blockage level 2 (IFSB2), combination of impeller fault and dry run (IFSB5), combination of impeller fault and discharge blockage level 1 (IFDB1), combination of impeller fault and discharge blockage level 5, pitted cover fault (PC), combination of pitted cover fault and suction blockage level 2 (PCSB2), combination of pitted cover fault and dry run (PCSB5), combination of pitted cover fault and discharge blockage level 1 (PCDB1), combination of pitted cover fault and discharge blockage level 5 (PCDB5).

Table 5. Results of classification for 25% AWGN data

Train/ Test Speed (Hz)	Fault classification accuracy															
	HP	SB2	SB5	DB1	DB5	IF	IFSB2	IFSB5	IFDB1	IFDB5	PC	PCSB2	PCSB5	PCDB 1	PCDB 5	C_a
30	**87.8**	**97.6**	**100**	**100**	**90.2**	**100**	**100**	**100**	**100**	**100**	**100**	**95.1**	**95.1**	**100**	**97.6**	**97.6**
35	100	100	100	100	100	100	100	100	100	100	100	100	100	87.8	100	99.2
40	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
45	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
50	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
55	**100**	**100**	**100**	**100**	**99.7**	**100**	**100**	**100**	**100**	**93.5**	**100**	**95.1**	**100**	**97.6**	**100**	**99.1**
60	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
65	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Healthy pump (HP), suction blockage level 2 (SB2), dry run (SB5), discharge blockage level 1 (DB1), discharge blockage level 5 (DB5), impeller fault (IF), combination of impeller fault and suction blockage level 2 (IFSB2), combination of impeller fault and dry run (IFSB5), combination of impeller fault and discharge blockage level 1 (IFDB1), combination of impeller fault and discharge blockage level 5, pitted cover fault (PC), combination of pitted cover fault and suction blockage level 2 (PCSB2), combination of pitted cover fault and dry run (PCSB5), combination of pitted cover fault and discharge blockage level 1 (PCDB1), combination of pitted cover fault and discharge blockage level 5 (PCDB5).

To get an understanding of the competitiveness of the present work, the results of it are compared with some of the other state-of-art works in this field. Table 6 presents the comparison. From the table, it can be seen that not many researchers have taken into consideration the combined effect of the mechanical and hydraulic faults of the CP. The works have majorly been confined to a single operating CP speed, but in the present work eight CP operating speeds are considered. From the results discussed, it can be clearly seen that the fault manifestation changes with the operating CP speed. Therefore, taking into account a range of CP operating speeds is important. Also, it can be seen that very few researchers have worked with corrupted signals to identify the CP faults. But, in reality data can be never immune of noise (or external disturbance). Hence, this consideration is important. The methodology developed in this paper is very robust and could classify 15 faults at 8 operating speeds giving an overall classification accuracy of more than 99%. Therefore, the features identified and the methodology developed has practical applicability.

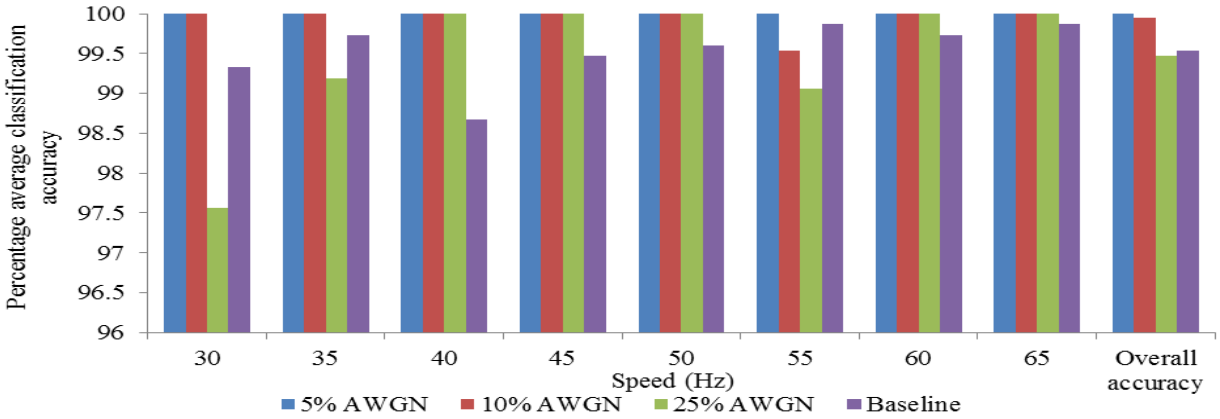

Figure 11. Comparison of results of classification for 5% AWGN, 10% AWGN, 25% AWGN and no-noise data. Additive white Gaussian noise (AWGN), healthy pump (HP), suction blockage level 2 (SB2), dry run (SB5), discharge blockage level 1 (DB1), discharge blockage level 5 (DB5), impeller fault (IF), combination of impeller fault and suction blockage level 2 (IFSB2), combination of impeller fault and dry run (IFSB5), combination of impeller fault and discharge blockage level 1 (IFDB1), combination of impeller fault and discharge blockage level 5, pitted cover fault (PC), combination of pitted cover fault and suction blockage level 2 (PCSB2), combination of pitted cover fault and dry run (PCSB5), combination of pitted cover fault and discharge blockage level 1 (PCDB1), combination of pitted cover fault and discharge blockage level 5 (PCDB5).

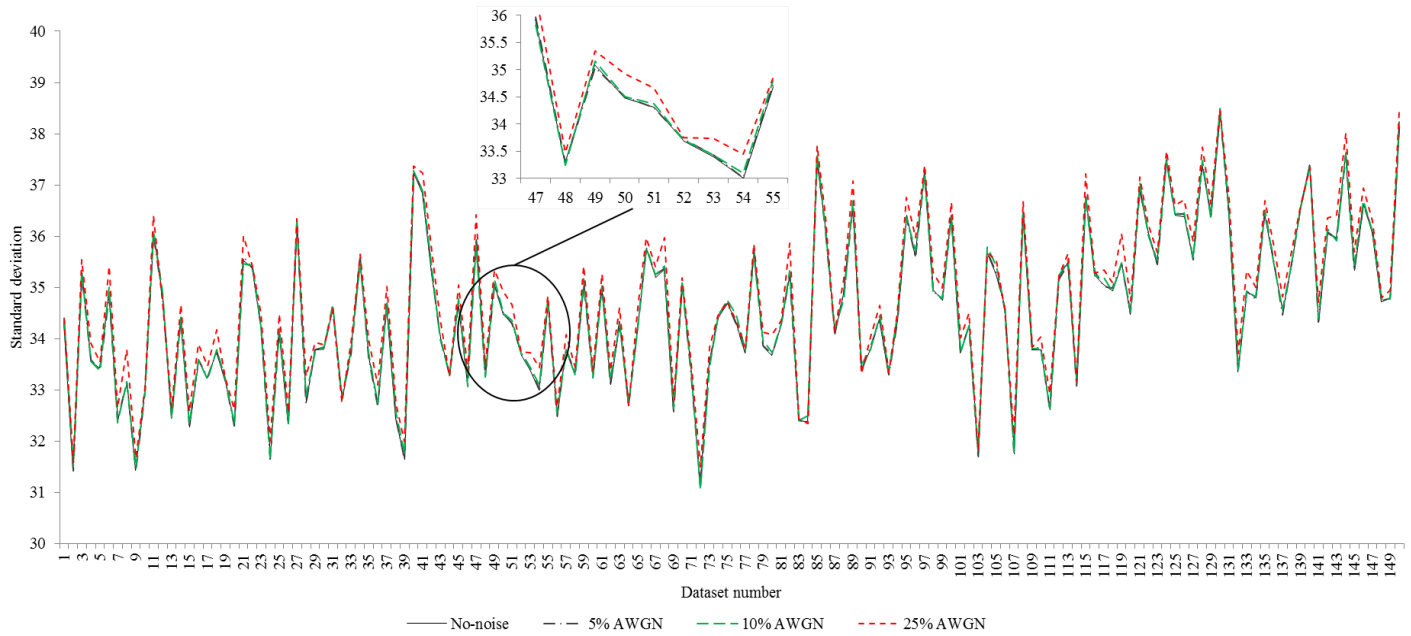

Figure 12. Comparison of standard deviation feature for no-noise, 5% AWGN, 10% AWGN, 25% AWGN for HP fault at 30 Hz speed.

Case study: Seawater lifting vertical pump

Problem statement: The pump and motor are submerged 35 meters below the platform within a caisson (as shown in Figure 12) and there is no instrumentation fitted (other than sporadic visual indication of the motor current). Even total pump flow at deck level is not measured and pressure is controlled using an overboard spill. The pump fails frequently due to the blockage because of debris (as shown in Figure 3). The design of the pumping system is such that the motor relies on the flow to provide for its cooling. The motor overheats causing damage to the windings. There is also pump pressure imbalance across its stages causing high thrust loads and failure of the shaft line, thrust bearing mounted within the motor. This allows the impeller to move axially, rub on the wear rings and then seize.

Table 6. Comparison of present work with other state-of-art work in the field

Paper/ Year	**Types of faults**	**Domain used**	**Features considered**	**Classifiers/ best prediction accuracy**	**Remarks**
(Sakthivel *et al.*, 2010)	Bearing faults, seal defects, impeller defects, combination of bearing and impeller faults and cavitation	Time - domain	Mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum and sum	Decision tree and rough set fuzzy/ 99.3% (decision tree fuzzy)	1. Vibration signals used 2. Algorithm tested at only one operating speed of the CP 3. Single speed characterization
(Azadeh *et al.*, 2010)	Two faults (no details of the faults mentioned)	Time-domain	Flow, pressure, velocity, temperature, vibration	SVM-GA, SVM-PSO, ANN/ 93.3% (SVM-PSO, SVM-GA, SVC) corrupted data	1. Used algorithms to perform binary data classification of corrupted data. 2. Single speed characterization
Rapur J.S and R. Tiwari (current work)	Suction blockages, Discharge blockages, impeller defects, impeller defects together with suction and discharge blockages, pitted cover plate faults, pitted cover plate together with suction and discharge blockages, dry run	Time-frequency domain	Standard deviation, reciprocal of standard deviation, RMS, RSS	SVM/ 99.5% (corrupt data), 99.9% (baseline data)	1. Vibration signals and motor line current used. 2. Combined mechanical and hydraulic faults considered 3. Corrupted data classified 4. Multi speed characterization (8 speeds)

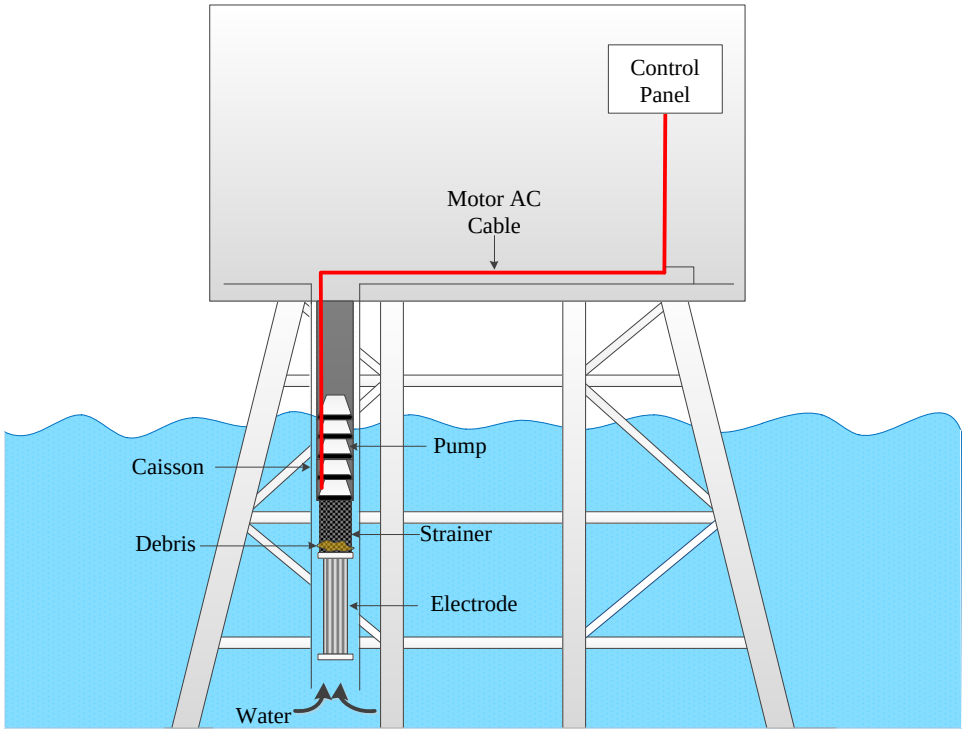

Figure 12. Schematic of seawater lift pump application

Solution: The pump failure can be avoided if initial stages of the pump blockage can be identified with high precision. The pump at present is not monitored with any instrumentation. The only available signal reading is that of the motor current. The motor current varies with the changing load on the CP. In this application, a sporadic visual inspection based method is used to identify anomalies in the pump operation. However, a visual inspection based method has a few drawbacks, including,

(1) Corruption of signal data which can cause variation in signal patterns, therefore appropriate signal filtering techniques need to be used,
(2) Highly experienced professionals are required to accurately identify the fault condition,
(3) It is highly involved to identify the faults at their formative stages, because they may not be able to give a clear indication of the fault characteristics.

Therefore, to avoid any mishaps the CP can be frequently monitored using the algorithm developed in this paper. This monitoring can be done at two levels.

1. When there is not enough fault history data available, the statistical parameters extracted from the motor current signal can be used to find the shift in their values, and thus attribute them to the anomalies. (Like shown in Figure 7)
2. When enough data is recorded, the data could be fed to the algorithm developed in this paper to find the faults. The motor line current signature is stored at different operating conditions of the CP and can be used to train the algorithm. The real-time current data may be then used to identify the commencement of the CP fault (in this case, a blockage fault). The data acquired from the CP in the faulty and non-faulty conditions needs to be given a class label. Once the class label is assigned, statistical features may be extracted from the faults. These features may be then used to train and test the SVM algorithm. This trained algorithm can be supplied with data from the CP at its current operating condition. Whenever the blockage starts the classifier gives an indication of the fault with almost 100% accuracy. This information can be supplied to the operator, so that necessary corrective actions may be initiated.

The advantages of using such a method are, (1) does not incur any additional cost to the industry as the motor line current is already being measured (economically viable), (2) identifies inception of fault so that remedial action may be promptly initiated and (3) helps avoid frequent failures of the pump systems. There is also scope of automatic maintenance by using this method.

This algorithm can also be applied in industries using corrosive/ erosive pumping fluids. Apart from the faults considered in this paper the prescribed methodology can be used to identify many other CP faults or faults in its ancillary components (like the bearings, seals, couplings etc.). Also, other signals like the noise spectrum, transient pressure signals, acoustic emission etc. can be used for fault

identification using the proposed methodology.

CONCLUSIONS

A unique flexible algorithm is proposed in this paper to classify multiple interdependent faults as well as independently existing faults on the CP using the CWT of vibration and motor current signals. A total of 33 faults are seeded on the CP including suction blockages (with varying severity), discharge blockages (with varying severity), impeller defects, impeller defects together with blockages, pitted cover plate faults, pitted cover plate faults together with blockages, dry run. The algorithm is tested for wide range of CP operations. Also, to account for the noise generated while handling the signals, the data is corrupted with 5%, 10% and 25% additive white Gaussian noise. The algorithm performed remarkably well in isolating various faults on the CP with great precision. In case the CP is directly driven by the motor, then it is expected that the performance would further improve with the line-current signals. The proposed methodology may be used by industry practitioners to detect various CP faults when fault history data is readily available. As a future scope, this algorithm may be tested with real-time industrial data. Also, the variance in manufacturing tolerances in pumps can be estimated and used as a noise, so that both the quality and reliability aspects of the CP can be considered simultaneously. Furthermore, a control loop may be designed so as to alarm the operator about the onset of fault so that immediate correction measure may be initiated.

NOMENCLATURE

A_N	Amplitude of the noise
A_S	Amplitude of the signal
C	Error penalty
C_a	Percentage classification accuracy
K	Kernel function
N	Additive white Gaussian noise
S	Noisy signal
SNR_{dB}	Signal to noise ratio in decibels
S_o	Original signal
$\mathbf{x}$	Data input vector
$\mathbf{w}$	Vector normal to hyper plane
ξ	Noise in slack variables
γ	RBF penalty parameter
σ	Standard deviation
σ_R	Reciprocal of standard deviation
RMS	Root mean square
RSS	Root sum of squares

ABBREVIATIONS

ANN	Artificial neural network
AWGN	Additive white Gaussian noise
CP	Centrifugal pump
CV	Cross validation
CWT	Continuous wavelet transform
DAG	Direct acyclic graph
DB	Discharge blockage
HP	Healthy pump
IF	Impeller fault
MFS^{TM}	Machine fault simulator
OAA	One against all
OAO	One against one
PC	Pitted cover plate fault
RBF	Radial basis function
SB	Suction blockage
SVM	Support vector machine

REFERENCES

Alabied, S., O. Hamomd, A. Daraz, F. Gu and A. Ball, Fault Diagnosis of Centrifugal Pumps based on the Intrinsic Time-scale Decomposition of Motor Current Signals. in Proceedings of the 23rd International Conference on Automation & Computing University of Huddersfield, 2017.

Azadeh, A., V. Ebrahimipour and P. Bavar, 2010: A fuzzy inference system for pump failure diagnosis to improve maintenance process: The case of a petrochemical industry. *Expert Systems with Applications,* **37,** 627-639.

Azadeh, A., M. Saberi, A. Kazem, V. Ebrahimipour, A. Nourmohammadzadeh and Z. Saberi, 2013: A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. *Applied Soft Computing,* **13,** 1478-1485.

Bordoloi, D. and R. Tiwari, 2017: Identification of suction flow blockages and casing cavitations in centrifugal pumps by optimal support vector machine techniques. *Journal of the Brazilian Society of Mechanical Sciences and Engineering***,** 1-12.

Chang, C.-C. and C.-J. Lin, 2011: LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.,* **2,** 1-27.

Chudina, M., 2003: Noise as an Indicator of Cavitation in a Centrifugal Pump. *Acoustical Physics,* **49,** 463-474.

Ebrahimi, E. and M. Javidan, 2017: Vibration-based classification of centrifugal pumps using support vector machine and discrete wavelet transform. *Journal of Vibroengineering,* **19,** 2586-2597.

Führ, H., 2005: *Abstract Harmonic Analysis of Continuous Wavelet Transforms*. Springer Berlin Heidelberg, Berlin, Germany.

Hart, R. J., Pumps and Their Systems-A Changing Industry. in Proceedings of the 19th International Pump Users Symposium and Short Courses, 2002, p. 141-144.

Hennecke, F., Reliability of Pumps in Chemical Industry. in Proceedings of the Pump Users International Forum, Karlsruhe Congress Centre, Karlsruhe, Germany, 2000.

Hsu, C.-W., C.-C. Chang and C.-J. Lin, 2003: A practical guide to support vector classification. 1-16.

Janani, S. R. and R. Tiwari, 2017: Experimental Time-domain Vibration Based Fault Diagnosis of Centrifugal Pumps using SVM. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering,* **3,** 044501-044507.

Kallesoe, C. S., V. Cocquempot and R. Izadi-Zamanabadi, 2006: Model based fault detection in a centrifugal pump application. *IEEE Transactions on Control Systems Technology,* **14,** 204-215.

Muralidharan, V. and V. Sugumaran, 2013: Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump. *Measurement,* **46,** 353-359.

Muralidharan, V., V. Sugumaran and V. Indira, 2014: Fault diagnosis of monoblock centrifugal pump using SVM. *Engineering Science and Technology, an International Journal,* **17,** 152-157.

Perovic, S., P. J. Unsworth and E. H. Higham, Fuzzy logic system to detect pump faults from motor current spectra. in Proceedings of the Industry Applications Conference, Thirty-Sixth IAS Annual Meeting Conference 2001, p. 274-280

Rapur, J. S. and R. Tiwari, A Compliant Algorithm to Diagnose Multiple Centrifugal Pump Faults With Corrupted Vibration and Current Signatures in Time-Domain. in Proceedings of the ASME 2017 Gas Turbine India Conference, 2017, p. V002T005A007-V002T005A007.

Rapur, J. S. and R. Tiwari, 2018: Automation of multi-fault diagnosing of centrifugal pumps using multi-class support vector machine with vibration and motor current signals in frequency domain. *Journal of the Brazilian Society of Mechanical Sciences and Engineering,* **40,** 278.

Sakthivel, N. R., B. B. Nair, M. Elangovan, V. Sugumaran and S. Saravanmurugan, 2014: Comparison of dimensionality reduction

techniques for the fault diagnosis of mono block centrifugal pump using vibration signals. *Engineering Science and Technology, an International Journal,* **17,** 30-38.

Sakthivel, N. R., V. Sugumaran and B. B. Nair, 2010: Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump. *Mechanical Systems and Signal Processing,* **24,** 1887-1906.

Sakthivel, N. R., V. Sugumaran and B. B. Nair, 2012: Automatic rule learning using roughset for fuzzy classifier in fault categorization of mono-block centrifugal pump. *Applied Soft Computing,* **12,** 196-203.

Tiwari, R., 2017: *Rotor Systems: Analysis and Identification*. CRC Press.

Vapnik, V., 1995: *The nature of statistical learning theory*. Springer science & business media.

Vapnik, V. N., 1999: An overview of statistical learning theory. *IEEE transactions on neural networks,* **10,** 988-999.

Venkatachalam, R., 2014: *Mechanical Vibrations*. PHI Learning.

Widodo, A. and B.-S. Yang, 2007: Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing,* **21,** 2560-2574.

Young, R. K., 2012: *Wavelet Theory and Its Applications*. Springer US.

ACKNOWLEDGEMENTS

The authors would like to thank the infrastructure and financial support provided by Indian Institute of Technology Guwahati for carrying out this research. The authors would like to recognize the LIBSVM tool, which was very useful in carrying out the present study. It is freely available at https://www.csie.ntu.edu.tw/~cjlin/libsvm/ (Chang & Lin, 2011).