



Automation of multi-fault diagnosing of centrifugal pumps using multi-class support vector machine with vibration and motor current signals in frequency domain

Janani Shruti Rapur¹ · Rajiv Tiwari¹

Received: 7 August 2017 / Accepted: 19 April 2018 / Published online: 5 May 2018
© The Brazilian Society of Mechanical Sciences and Engineering 2018

Abstract

Reliable detection and isolation of centrifugal pump (CP) faults is a challenging and important task in the modern industries. Hence, this paper proposes an artificial intelligence-based multi-fault detection of CPs driven by induction motor. The intelligent fault detection methodology is developed based on the multi-class support vector machine (MSVM). The mechanical and hydraulic faults in CPs are mutually dependent and therefore may exist concurrently. Hence, in the present research, an assortment of various flow instabilities like the suction re-circulation, discharge re-circulation, pseudo-re-circulation and dry runs are considered coexisting with mechanical faults, like the impeller cracks and pitted cover plate faults. The power spectrum of the CP vibration and the induction motor line-current data is used for monitoring the CP condition. The best statistical feature combination is selected based on a wrapper model. Gaussian radial basis function (RBF) is used for the kernel mapping. In addition, the RBF kernel parameter (width) and MSVM parameters are optimally selected using a fivefold cross-validation technique. Also, variation of operating speed of the CP drastically changes the system vibration level owing to the change in fault manifestations; hence, in the present work a methodology that is independent of CP operation is proposed and tested. Thereafter, it is observed that the proposed methodology is remarkably robust and successfully classifies multiple individual as well as coexisting CP faults at all the tested CP speeds.

Keywords Centrifugal pump · Multi-fault diagnosis · Power spectrum · Multi-class support vector machine (MSVM)

1 Introduction

In the modern industries, fault diagnosis of the mechanical equipment is being given great importance [1, 2]. Centrifugal pumps (CPs) are preferred hydraulic turbomachines for industries with applications ranging from food processing industries to oil refineries. They are very robust in design and can cater to versatile flow requirements. It is estimated that a typical chemical plant operates at an average of one CP per employee [3]. Also, 20% of the energy produced globally is said to be consumed in running CPs [4]. They thus form vital components to sustain the process course of the plant. Thus, active condition

monitoring of CPs is important to ensure increased availability of the plant and operators' safety.

Faults in CPs can be broadly classified as mechanical faults and fluid-flow-induced faults. Interestingly, these faults are dependent on each other, which means that one CP fault type can stimulate the occurrence of the other [5]. (More details on the interdependence of mechanical and hydraulic faults are given in Sect. 5.) Therefore, it is not fitting to consider the hydraulic and mechanical CP faults independent of each other. Hence, to take the combined effects of these into account, the faults listed in Table 1 need to be considered. Here, re-circulation is defined as the turnaround of fluid-flow patterns at the inlet or outlet tips of the impeller blades. This re-entering flow results in the formation of vortex on the pressure side of the impeller blade. In case the flow possesses enough energy, the fluid re-circulation results in the material damage [6].

In Table 1, it is evident that different flow instabilities have distinct causes and effects on the CP system and its internal components. Thus, the combined effects of these

Technical Editor: Kátia Lucchesi Cavalca Dedini.

✉ Rajiv Tiwari
rtiwari@iitg.ernet.in

¹ Department of Mechanical Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, India

Table 1 CP fault names, their causes and effects

S. no.	Fault name	Fault cause	Fault effect
1.	Suction blockage [5, 12]	Obstruction in the inlet pipe of the CP due to unclean liquids/surface damages resulting in reduction in cross section of suction	Discharge re-circulation [42], cavitation damage on (a) the pressure side of the vane at the discharge of impeller, axial movement of shaft and (b) inlet of diffuser vanes of casing, cracking and failure of impeller shrouds, shaft failures, noise and excessive vibrations
2.	Discharge blockage [15]	Flow rates below BEP, closing of throttle valve	Suction re-circulation [42], cavitation damage on the pressure side of the vane at the inlet of impeller and, random crackling sound in the suction, surging in the suction of pump, noise and vibrations
3.	Impeller defects [7, 9–11, 15]	Casting errors of the impeller, cavitation damage, material corrosion, fatigue	Unbalance, pseudo-flow re-circulation, vibrations and noise [5]
4.	Pitted cover plate	Manufacturing defects, damages due to cavitation erosion and material corrosion	Turbulence to the flow pattern
5.	Dry run	Insufficient priming of the CP	Bearing, seal damages, vibrations and noise

faults may be catastrophic. Bubbles formed due to any of the aforementioned reasons not only decrease the head performance of the CP (because the CP has to invest some of its energy on compressing the air bubbles) but also may result in a cavitation-like damage on the internal components of CP.

Many researchers have worked on the CP fault diagnosis. Most of the work emphasized on the identification of mechanical faults of CP, like bearing faults, seal defects and impeller faults. Very limited work has been performed on the hydraulic CP fault diagnosis and/or combination of various CP faults. Muralidharan et al. [7] researched on detecting cavitation faults, bearing faults, impeller faults and both the bearing and impeller faults together. Sakthivel et al. [8, 9] diagnosed CP bearing faults, seal defects, impeller defects, cavitation and the bearing and impeller defects together. Wolfram et al. [10] presented a model-based approach to detect leakage faults, varying CP obstructions, cavitation, bearing faults and impeller defects. Wang et al. [11] investigated on a method to detect the bearing and impeller wears. In all the aforementioned studies, the focus majorly was on the mechanical CP faults. However, since mechanical faults are directly related to operating speeds of the CP, they can be picked out much more efficiently than the complex-flow-induced faults.

Some researchers have worked on identifying different flow-related faults in the CPs. Bordoloi and Tiwari [12] worked on identifying different levels of suction blockages on the CP. They reported an average classification accuracy of 56% at low speed of CP operation. In addition, light suction flow restriction could be identified with only 10% accuracy at low CP speeds. The possible reasons for such low prediction performance could be the (1) inappropriate

choice of the statistical features and/or (2) insufficient resolution of the sampled experimental signals. Chudina [13] and Alfayez [14] developed methods to detect the incipient cavitation. In these papers, the flow instabilities were diagnosed. However, the researchers have not considered any combined effects of mechanical faults with flow-related faults.

Perovic [15] took this aspect into consideration and tried to develop a technique to classify cavitation faults, discharge blockage faults, impeller defects and both discharge blockage and impeller defects. The discharge blockage combined with the impeller defects and the cavitation were classified at a rate of 83.7%, and the blockage and the cavitation were classified at a rate of 61%. Rapur and Tiwari [5] attempted classification of CP with suction blockage faults and impeller defects, but the results suggested the inability of the algorithm to identify faults at small suction restrictions, especially at low speeds. This was attributed to the similarity of the fault signatures of low levels of suction restriction of CP to that of the healthy CP.

Concisely, little emphasis has been given to the identification of coexisting mechanical and flow-induced faults of CPs. The classification rates obtained were very low while considering coexisting mechanical and hydraulic faults [5, 15]. Moreover, the algorithms developed for fault identification and diagnosis of CPs were limited to a single CP operating speed. But, in majority of the cases, CP operation is not confined to a single speed, and in complex nonlinear systems, it is a great challenge to develop accurate process models valid over wide operating range [10].

The prediction of defects in mechanical systems can be done either by the qualitative and statistical analysis of collected fault data or by developing mathematical models of the system or by machine learning techniques. In a qualitative analysis, the characteristic fault frequency deviations are studied to identify different faults [16]. However, there is a significant amount of human error involved in it. Some researchers have diagnosed CP faults using modeling techniques [10, 17]. However, it is very complex to accurately model the interdependent faults and identify them. With the advancement of modern computational capabilities, online machine learning and condition monitoring techniques are being preferred. There are numerous popular machine learning techniques including artificial neural networks (ANNs), fuzzy logic, empirical mode decomposition (EMD), decision tree, support vector machine (SVM), deep learning. Researchers have used various aforementioned machine learning techniques to diagnose faults in various mechanical systems including bearings, gears, induction motors and pumps [11, 18–28]. Some recent researches that have been conducted on the CP fault diagnosis are [5, 7–9, 12, 15, 22, 29]. Of all the above-mentioned methods, ANN is most popular among researchers. It utilizes the empirical risk minimization principle and works fine with large data. But the basic ANN suffers from local minima traps and complication in establishing hidden layer sizes and the learning rate [30]. On the other hand, SVM employs structural risk minimization (SRM) principle and has a clear mathematical formulation. Therefore, it is expected to give better learning performance as it can arrive at the global minimum [30, 31]. Taking into account the advantages offered by SVM, in the present work it is employed for machine learning and condition monitoring of the CP.

The choice of signal used to monitor the CP condition is very crucial. Popular signal choices are CP vibration [5, 9, 32, 33], motor line current [34, 35] and acoustic emission signals [14, 36]. The change in CP system operation due to faults also changes the load on the motor, and therefore, the motor line-current changes. Hence, the motor current has also been used by researchers to diagnose CP faults [15]. Vibration signatures are very resourceful, are sensitive to the fault conditions and are easy to acquire, thus making them the most preferred signals for the CP fault diagnosis [37]. Taking into credit the benefits offered by the vibration signals and also the motor line-current signals, in this work their combination would be used for CP fault diagnosis. Researchers have not attempted this previously.

The signal acquired can be analyzed in time domain, frequency domain and time–frequency domain. Time domain is sensitive and provides physical understanding of the signal, and thus, many researchers use this domain for

the fault diagnosis. However, there are three major advantages offered by spectral/frequency analysis over that of the time-domain analysis. The first is that it simplifies the comprehension of the waveform. Secondly, the physical property of the signal often depends on the frequency; thirdly, it is a mathematical tool for solving equations. Some papers on the fault diagnosis of CP in frequency domain are [7, 38]. Consequently, this research uses the frequency domain of the acquired signals for fault diagnosis. The acquired signal contains a lot of redundant information and is of high dimensionality; therefore, it cannot be directly fed to the machine learning algorithm. Hence, suitable statistical features need to be extracted from the raw data. In this paper, the wrapper model is used to select the best-performing statistical features.

The major contributions of the present work are:

- This paper considers the interdependence of mechanical and hydraulic CP faults. Attempt to classify mechanical and hydraulic faults in the CP existing both independently and in combinations is presented. Section 2 explains the fault dependency, and results of the classification are presented in Sect. 5.
- To take a suitable corrective action, it is imperative to identify the fault severity. However, not a lot of research has been done on identifying varying flow instabilities of CPs. In this paper, an attempt is made to isolate various severity levels of the suction and discharge blockages of CPs. Section 2 gives details of the suction and discharge blockages. Section 5 gives results of the classification.
- In this paper, a robust methodology, which works well at a wide range of CP operations, is proposed to be developed. In addition, it is attempted to diagnose the faults at intermediate test speeds. Section 3 gives the details of the methodology. The results are tabulated in Sect. 5.1(e).
- The sampling rate of data acquisition plays a significant effect on the classification performance obtained. A comparative study of the classification performance obtained with a high-frequency-resolution sampling and a low-frequency-resolution sampling is presented in Sect. 4.
- The CP fault diagnosis traditionally used vibration signals or acoustic signals, or motor line-current signals. The effectiveness of using the combination of the vibration and motor line-current signals is substantiated in this paper in Sect. 5.
- In this research, an attempt has been made to develop a methodology to classify thirty-three critical CP faults using SVM as the artificial intelligence technique. The statistical features/feature combinations that achieve

the best classification accuracy are identified, and their performance is demonstrated in Sect. 4.

The rest of this paper is arranged as follows: Section 2 describes the CP fault analysis test setup. A theoretical understanding of the proposed methodology is given in Sect. 3. Section 4 describes feature selection and effect of signal sampling rate on the fault prediction. Classification results are presented in Sect. 5. The last section gives the conclusions and future work.

2 CP fault analysis test setup

To perform the fault diagnosis, the faults are artificially seeded on two geometrically similar Oberdorfer 60P pumps. The CPs are made to have three operating configurations, one with the healthy impeller and the healthy cover plate, the second with the faulty impeller and the healthy cover plate and the third with the healthy impeller and the faulty cover plate. A total of 33 faults are considered on the CPs, which are broadly classified into ten fault families, which are (I) healthy pump, (II) suction blockage faults, (III) discharge blockage faults, (IV) impeller defects, (V) combination of impeller defects and suction blockages, (VI) combination of impeller defects and discharge blockages, (VII) pitted cover plate faults, (VIII) combination of pitted cover plate faults and suction blockages, (IX) combination of pitted cover plate faults and discharge blockages and (X) dry-run faults for all the CP configurations.

Furthermore, the suction and discharge blockages are given on the inlet and outlet of the CP, respectively, using mechanical modulating valves, as shown in Fig. 1a. The valves are calibrated to give the desired amount of flow restriction, i.e., 0% (B0) (full flow), 16.7% (B1), 33.3% (B2), 50% (B3), 66.6% (B4) and 83.3% (B5).

Further, the faults on the impeller are seeded by cutting through-through notches (2 per vane) on the impeller vanes, as shown in Fig. 1b. The notches are dissimilar and are asymmetrically placed. Subsequently, faults are also given on the cover plate. Pits are artificially created on the cover plate as shown in Fig. 1c. The nomenclature and descriptions of faults considered are given in Table 2. In the table, the *tick* (✓) mark denoted the presence of a fault and the *cross* (×) mark denoted the absence of it. The number alongside every fault abbreviation indicates the amount of flow restriction, i.e., 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 has been considered as a ‘dry-run’ condition. Here, IF stands for impeller faults; PC stands for pitted cover plate faults.

The CP under study is installed on a machine fault simulator (MFSTM) setup, as shown in Fig. 2. The schematic diagram of the experimental setup is given in Fig. 3. CP is driven by a 3- ϕ induction motor (4-pole, 1.6 kVA, 4.2 A, 0–240 V output voltage), using a belt-pulley drive. The vibration and current signatures from the CP and the motor are measured using two triaxial accelerometers and three line-current probes, respectively. Accelerometers are mounted on the CP casing and bearing housing locations. The CP is operated using a variable speed motor from 30 to 65 Hz in steps of 5 Hz. A wide speed range of the CP operation is chosen to check whether the developed algorithm shows any speed dependency on the classification performance. The data are sampled at two different sampling rates—20,000 samples/s (2000 samples/dataset) and 5000 samples/s (5000 samples/dataset). The details of the amount of data collected for each fault and for all the faults are specified in Table 3. Table 4 shows the data resolution obtained in the frequency domain with each of these sampling rates. It can be observed that both the sampling

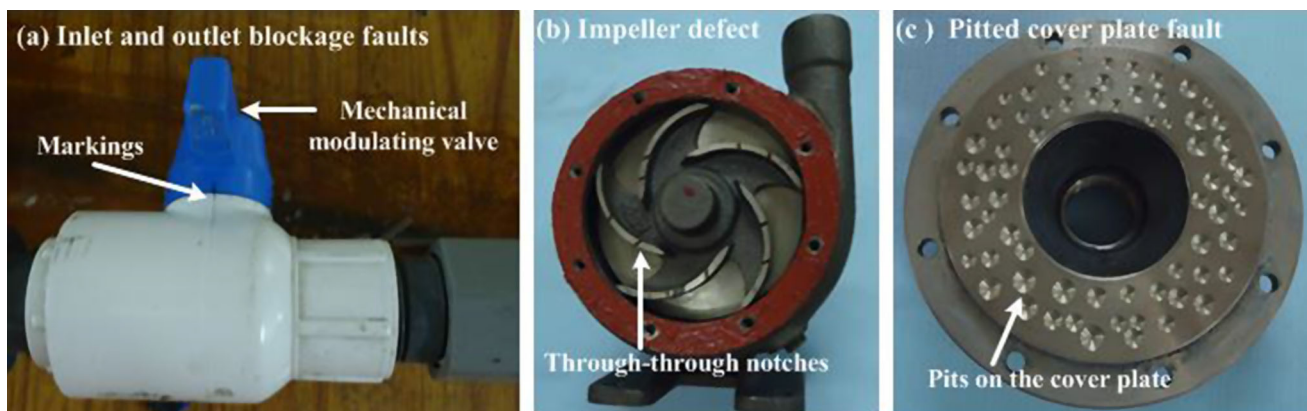
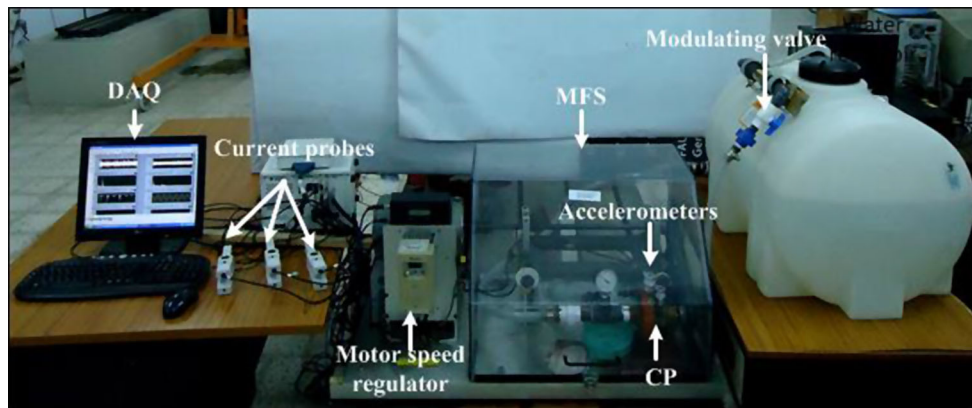


Fig. 1 a Mechanical modulating valve with markings; b cracks on the impeller; c pitted cover plate

Table 2 CP faults considered, their nomenclature and description of what they include

Pump configuration	Fault considered	Fault abbreviation	Suction blockage	Discharge blockage	Impeller defect	Pitted cover plate
I	Healthy CP	HP	×	×	×	×
II	Only suction blockages	SB i *	✓	×	×	×
III	Only discharge blockages	DB i *	×	✓	×	×
IV	CP with impeller defect	IFSB0	×	×	✓	×
V	CP with both impeller defect and suction blockages	IFSB i *	✓	×	✓	×
VI	CP with both impeller defect and discharge blockages	IFDB i *	×	✓	✓	×
VII	CP with pitted cover plate fault	PCSB0	×	×	×	✓
VIII	CP with both pitted cover plate fault and suction blockages	PCSB i *	✓	×	×	✓
IX	CP with both pitted cover plate fault and discharge blockages	PCDB i *	×	✓	×	✓

* $i = 1, 2, 3, 4, 5$

**Fig. 2** Experimental setup of a centrifugal pump with an indication motor

rates chosen show different Nyquist frequencies and data resolutions. This is so chosen to check the reliance of algorithm's classification performance on the resolution of data acquisition. The power spectral density of the raw data in linear scale, obtained from both high- and low-frequency-resolution data sampling, is plotted in Fig. 4a, b, respectively. From the figure, it can be seen that the frequency activity is typically within 1000 Hz for the system. Over that, the high-frequency-resolution sample (Fig. 4a) tends to carry more details (intricate frequency amplitude information) of the characteristic fault frequencies, whereas with the low-frequency-resolution data sampling (Fig. 4b) data of higher frequencies are obtained but with a loss of amplitudes at refined characteristic frequencies. That means, it is safe to say that instead of capturing data at higher frequencies (where there is not much vibration activity), a better resolution may be chosen to capture intricate frequencies.

The following observations were made during experimentation.

- The suction CP blockage up to SB2 level does not seem to change the flow pattern in CP. However, there is bubble formation observed between the SB2 and SB3 levels. The severity of bubble formation and the size of bubbles keep increasing with the increase in the suction blockage severity.
- The presence of impeller defects (even without any suction/discharge blockage) causes bubble formation.
- Pits on the cover plate also seem to disturb the flow pattern. Tiny bubbles are observed even in PCSB0 at increased speeds of the CP operation.
- Dry-run faults cause a lot of vibration and associated noise.
- The presence of discharge blockage reduces the pump discharge significantly and increases the vibrations.

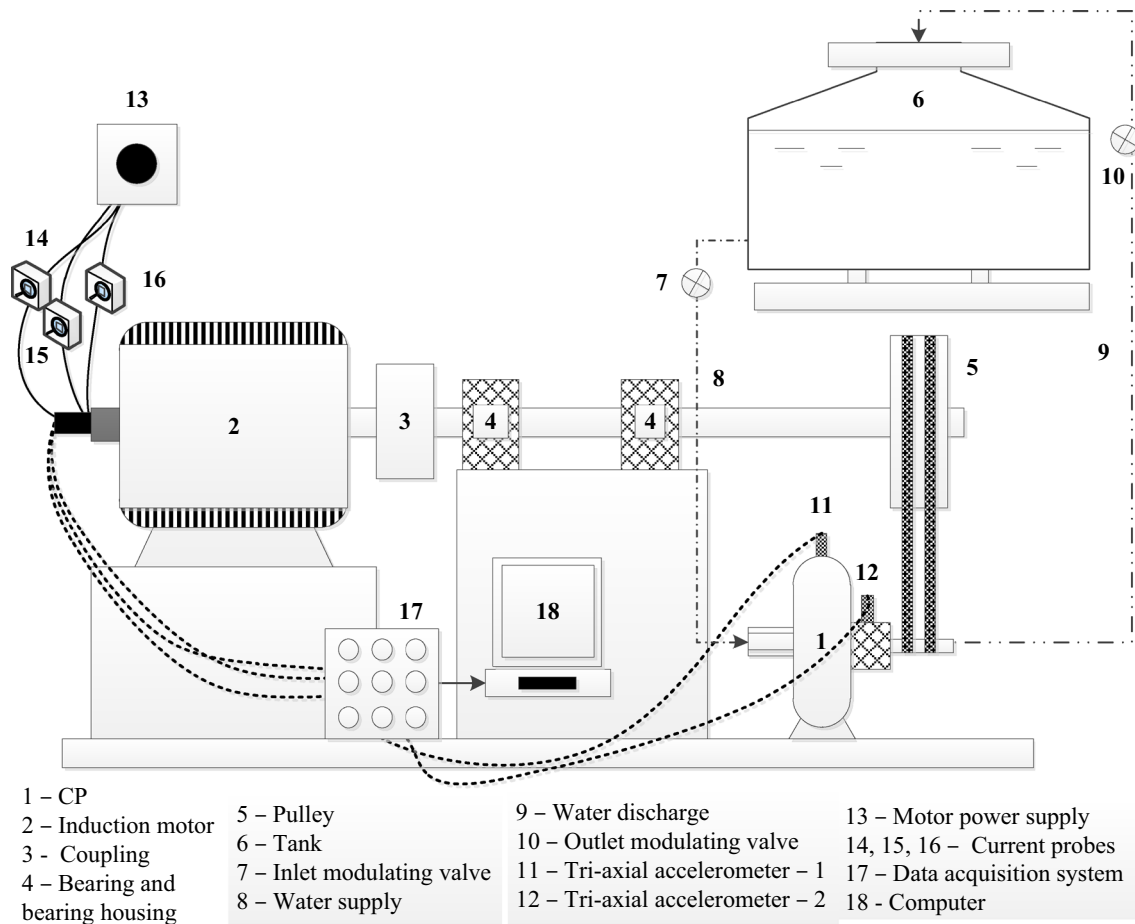


Fig. 3 Schematic diagram of the experimental setup

Table 3 Total data collected using different sampling rates of data acquisition

Sampling rate (kHz)	No. of samples/dataset in time domain	No. of samples/dataset in frequency domain (a)	No. of datasets (b)	No. of measurements using two triaxial accelerometers (c)	No. of measurements using three line-current probes (d)	No. of faults (e)	Total data collected per fault $f = \{a \times b \times (c + d)\}$	Total data collected for all faults $g = (f \times e)$
20	2000	1000	350	6	3	33	3,150,000	103,950,000
5	5000	2500	150	6	3	33	3,375,000	111,375,000

Table 4 Details of data resolution and frequency content in data acquired with different sampling rates

Samples to read per dataset (S)	2000	5000
Sampling rate	20,000 Hz	5000 Hz
The time taking for acquiring one dataset (T)	$2000/20,000 = 0.1$ s	$5000/5000 = 1$ s
Time taken to acquire single data point (ΔT)	$\frac{T}{S} = \frac{0.1}{2000} = \frac{1}{20000}$	$\Delta T = \frac{T}{S} = \frac{1}{5000}$
Resolution in frequency domain	$\frac{1}{T} = \frac{1}{0.1} = 10$ Hz	$\frac{1}{T} = \frac{1}{1} = 1$ Hz
Nyquist frequency (N, Hz)	$\frac{1}{2\Delta T} = 10000$ Hz	$\frac{1}{2\Delta T} = 2500$ Hz
Total time (s)	35	150
Total datasets	350	150
Data resolution in frequency	Low	High

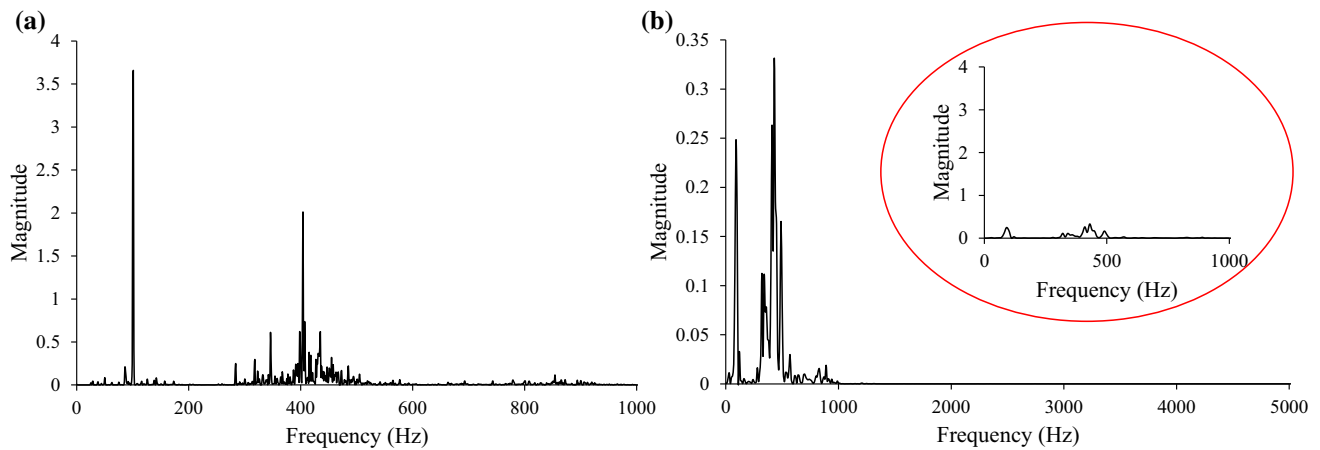


Fig. 4 Raw power spectral density data in linear scale from the accelerometer-1 for HP fault at 30 Hz with **a** high-frequency-resolution data sampling and **b** low-frequency-resolution data sampling. The y-axis has the units of $((\text{mm/s}^2)/\text{Hz})$

2.1 Interdependence of faults

The mechanical and hydraulic CP faults are not independent of each other. In this section, their symbiosis is explained.

As mentioned in the experimental observations, the CP with impeller cracks revealed bubble formation even in the absence of blockage. This implies that the impeller cracks act as flow instability inducers. The CP as shown in Fig. 5 runs in the backward curved vane configuration. Figure 5a shows the flow pattern inside a CP without impeller cracks, and Fig. 5b shows the flow pattern inside a CP with impeller cracks. The convex side of the impeller blade is on a higher pressure than the concave side. In Fig. 5b, the HPZ stands for high-pressure zone and the LPZ stands for low-pressure zone. This pressure difference makes the fluid

to flow through the cracks as shown in Fig. 5b. Hence, some fluid, which is supposed to flow along the convex vane face, leaks into the LPZ of the following vane as shown in Fig. 5b. This flow disrupts the flow pattern of the following impeller blade and may cause a *pseudo-re-circulation* [5]. Re-circulation is always accompanied by low-pressure pulsation zones, and hence, this might be the reason for bubble formation

Any bubble formation inside the CP system results in a decreased CP performance and material erosion on internal CP surfaces. The collapse of bubble cloud creates a *shock wave*. This wave travels in the fluid, and its magnitude attenuates as it gets closer to the solid surface. Near the solid surface solitary bubbles are present. When the shock wave reaches the bubbles, they oscillate and a micro-jet phenomenon may occur, as shown in Fig. 6. When this

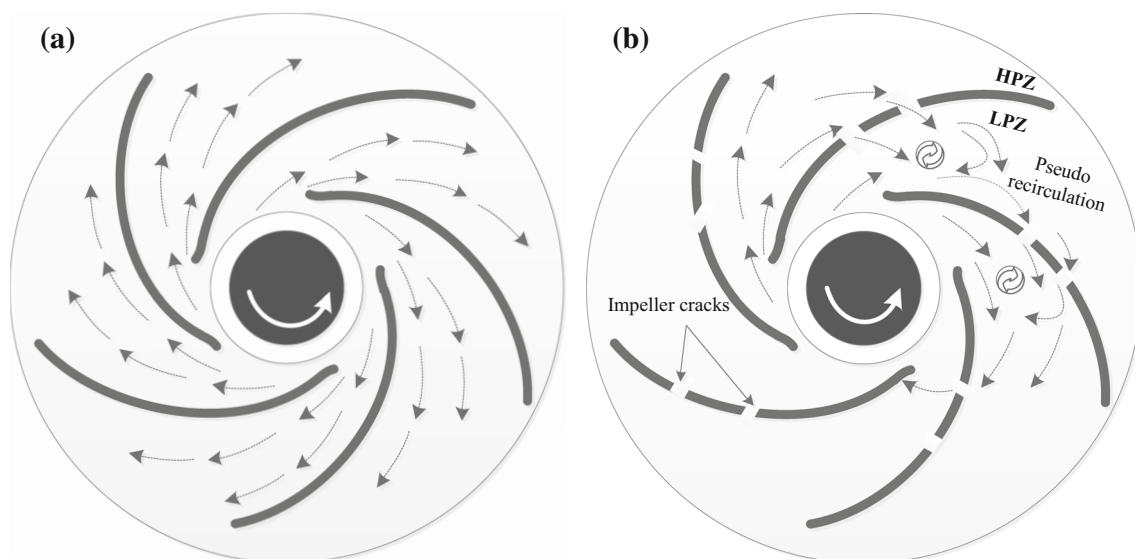


Fig. 5 **a** Flow pattern over healthy impeller blades; **b** flow pattern over cracked impeller blades

Fig. 6 **a** Pressure wave emission; **b** micro-jet formation; **c** pit formation and material erosion

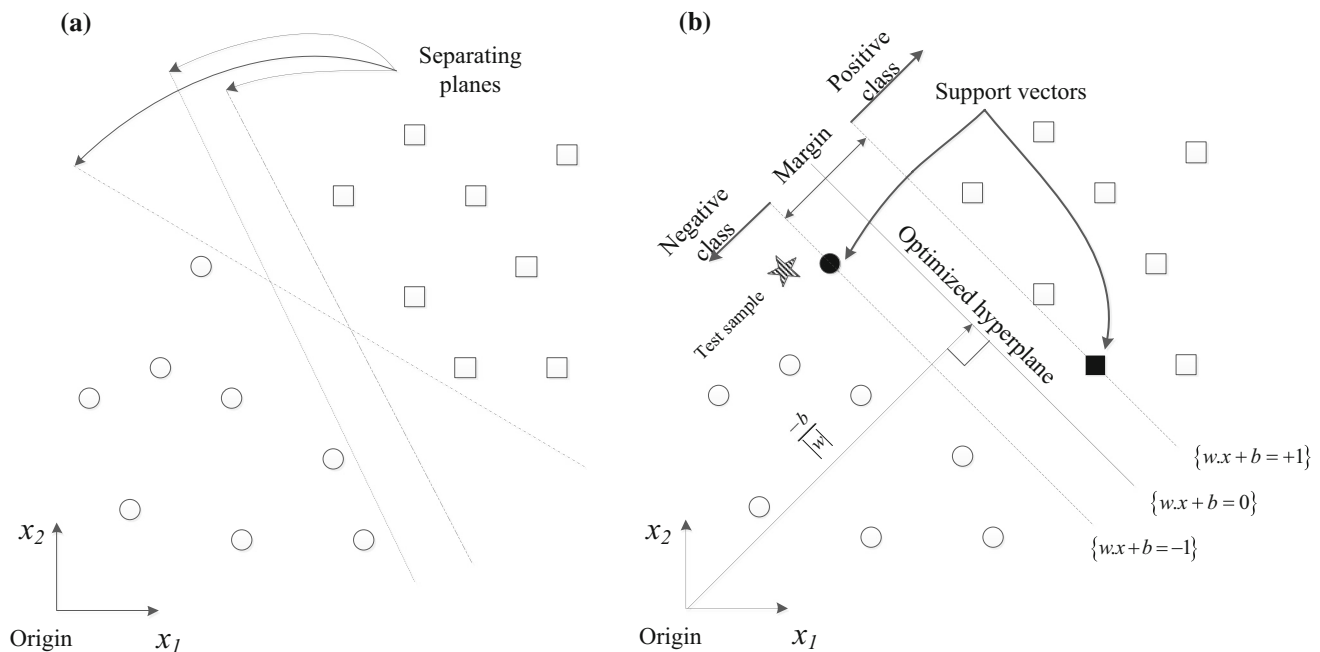
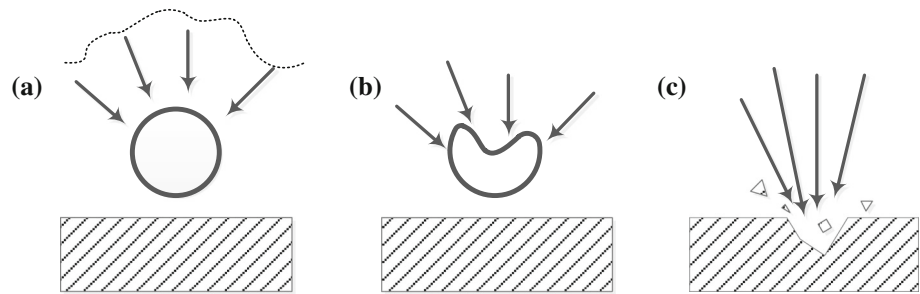


Fig. 7 **a** Various orientations of the data separating planes; **b** the optimal orientation of the SVM hyperplane

high-velocity micro-jet impinges on the solid surface, pitting takes place [39]. Material erosion on the impeller blades or any internal surface of the CP causes or enhances the mechanical faults. Therefore, the mechanical and hydraulic faults form a vicious fault cycle and hence need to be curbed at their formative stages itself.

To sum it up, the observations suggest that the re-circulation effects in the CP without impeller defects/pitted cover plate faults are due to the suction flow restriction and discharge flow restriction governed by the blockage. While in a pump with impeller defects, the cracks govern a pseudo-re-circulation. It is also expected that the pits on the cover plate cause a flow pattern disruption in the CP.

3 CP fault diagnosis methodology

As discussed in the literature review, due to the fair share of advantages offered by the SVM, it would be adapted as the artificial intelligence technique for the fault diagnosis

of the CP. A brief description of the working of SVM is presented in this section. However, for greater understanding of SVMs the following works may be referred [30, 31].

The SVM is a supervised learning technique. Fundamentally, SVM is a binary data classifier. From the data that need to be classified, suitable features are extracted. Let \mathbf{x} represent a feature vector of dimension N extracted from the data. For every vector \mathbf{x} , a class label y is allotted based on which data category it represents (positive class or negative class). Consider a binary classification case to classify p sets of training data $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_p, y_p)$ into two classes, where $\mathbf{x}_i \in \mathbb{R}^N$, $i = 1, 2, \dots, p$, and $y_i \in \{-1, +1\}$ specifies its class label. If we presume that the two classes of data are linearly classifiable by a hyperplane described by $\mathbf{w} \cdot \mathbf{x} + b = 0$ (where symbol \cdot is the dot product) in a space H , then the optimal hyperplane is that which maximizes the margin defined by *support vectors*. Let the two classes of data be represented by hollow circles and hollow squares, as shown in Fig. 7a. There can be

infinite orientations of planes separating the two classes of training data. But, only a hyperplane, which maximizes the *margin* between the support vectors denoted by dark markers, is chosen, as shown in Fig. 7b. The margin is defined as the Euclidean gap between the sorting hyperplane and the support vector of every class. The hyperplane is found by solving the following optimization problem,

$$\begin{aligned} & \text{Minimize} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^p \xi_i \\ & \text{Subject to} \quad y_i [\mathbf{w}^T \cdot \mathbf{x}_i + b] \geq 1 - \xi_i \\ & \quad \quad \quad \xi_i \geq 0, \quad i = 1, 2, \dots, p \end{aligned} \quad (1)$$

where ξ_i is the slack variable to account for incorrect classification, w is the weight vector, b is the bias and C is cost function which defines a bargain among model intricacy and learning error. In many cases where the data are not linearly classifiable, the data are transformed from the input space to a higher-dimensional feature space using a kernel function. In this transformed space, the data become linearly separable. Kernel function is given by $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$. The choice of kernel function plays a very important role in the accuracy of fault identification. From the literature, it has been found that Gaussian RBF kernel is very versatile [40]. Therefore, in the present work, it is proposed to adopt the Gaussian RBF kernel function. It is defined as,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \quad (2)$$

By introducing a kernel function, the optimization problem can be written in dual form using Lagrange multipliers α_i ($i = 1, 2 \dots p$), as

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^p \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \quad (3)$$

The algorithm is very sensitive to the kernel parameter γ and the SVM penalty parameter, C . The hyperplane, while using Gaussian RBF parameter, is an arrangement of bell-shaped curves placed at the support vectors. The width of each bell-shaped curve is inversely proportional to the value of γ . If the width of the bell-shaped curve is smaller than the minimum distance between the support vectors, the classification leads to the over-fitting of the data and does not capture the tendency of the data. In the other case, where the width of the bell-shaped curve is greater than the maximum distance between the support vectors, it leads to under-fitting. Cost function C marks a trade-off between the simplicity of the classification curve and the misclassification of the training sample. With the increase in the value of C , the complexity of the curve also increases. If γ were very high, any amount of fine-tuning of C would not

help in improving the performance of the classification algorithm. Hence, in this paper the best combination of γ and C is selected using a grid search and cross-validation (CV) technique. In this technique, the training data are divided equally into n subsets. The algorithm is first trained with $(n - 1)$ subsets from these n , and one subset is used for the testing. This process is repeated such that all the n subsets test one at a time. The γ and C parameters that give the best *cross-validation accuracy* (CV accuracy) are retained. These parameters are then used to train the optimized SVM. The accuracy of SVM is defined as the best average classification accuracy obtained using the selected γ and C parameters.

It is very unlikely that there would only be two classes of data that need to be classified. However, SVM is primarily a binary data classifier. Hence, to enable SVM to deal with more than two classes of data many models were proposed in the literature, including one-against-one (OAO) approach, one-against-all (OAA) approach and direct acyclic graph (DAG) approach. In the present work, to account for the multi-fault classifications, the OAO is employed to make the SVM a multi-class data classifier or multi-support vector machine (MSVM). In OAO for classifying p categories of data, $p(p - 1)/2$ binary classifiers are created. Each one trains data from two different classes. A voting strategy is used to classify the data while testing. While testing sample x , if a binary classifier says that the sample belongs to i th class, then the vote for class i is incremented by one. Lastly, sample x is classified into that particular class, which carries the maximum votes for it.

A flowchart of the present methodology is given in Fig. 8. The proposed methodology consists of the following seven steps:

1. The data are acquired from the CP using the accelerometers and line-current probes.
2. These data are in time domain and are converted to its power spectrum.
3. Suitable features are extracted from the data, and best features are selected using the wrapper model.
4. A subset of the acquired data is used for training the MSVM algorithm. The remaining data are kept aside and are used for testing the algorithm.
5. Cross-validation technique is used to select the best MSVM parameters.
6. The final trained algorithm with the optimized MSVM parameters is then tested with the remaining subset of the test data.
7. Faults are segregated into their respective groups, and the performance of the classifier is given by the percentage classification accuracy.

As a performance measure, the classification accuracy of the algorithm is calculated. It is given by,

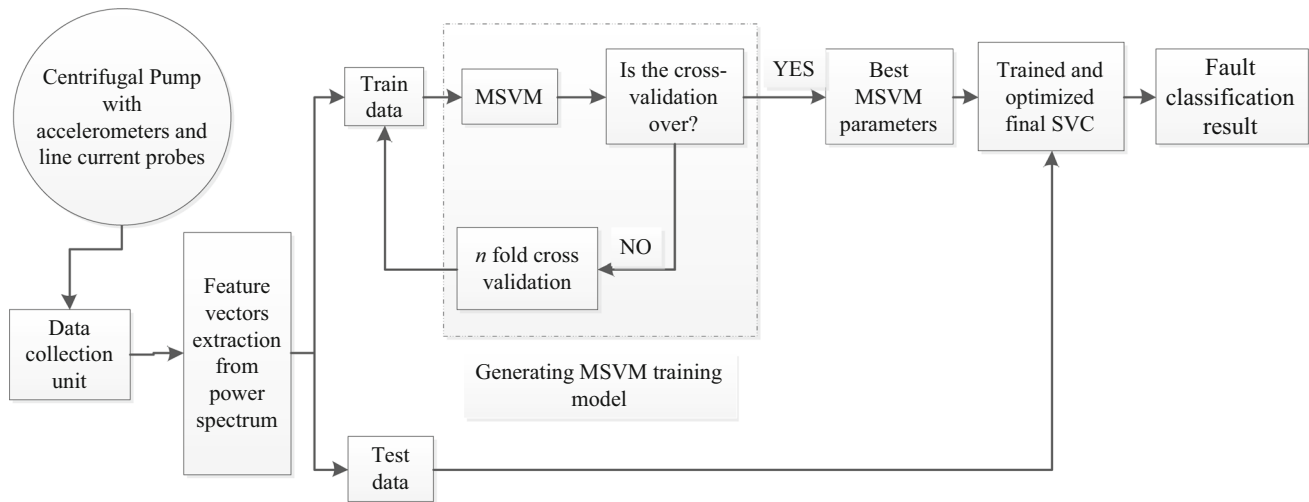


Fig. 8 Flowchart of the classification methodology

$$\text{Classification accuracy} = \frac{\text{Number of correctly predicted data}}{\text{Total number of testing data}} \times 100\% \quad (4)$$

Apart from the classification accuracy, in the case of a multi-class classification, the average classification accuracy is defined as the average of classification accuracies obtained for each fault. In addition, another term called the *overall classification accuracy* is defined, which is the average of average classification accuracy over the entire speed range. The overall classification accuracy gives a bird's eye view of the classifier's performance at all the CP operating speeds.

The LIBSVM [41] version 3.20 is used as the SVM tool. This package is tried and tested by many researchers, and it is proved very effective.

4 Feature selection and effect of the sampling rate on the fault predictions

Figure 9 shows the power spectrum of measured data that are plotted for various CP faults at 30 Hz speed using the data from accelerometer 1, accelerometer 2 and line-current probes. As it can be seen from the figure, these data are of very high dimensionality and contain redundant information. Therefore, if these data are directly used to train the MSVM classifier, the classification results may not be satisfactory. Hence, proper statistical features need to be selected which carry the fault information and are sensitive to the variation of fault severities. There are many methods for selection of the best features, namely principal component analysis (PCA), genetic algorithm (GA) and

wrapper model. In the present work, the wrapper model is used for the feature selection. Using it, the best features are selected based on the classification performance they demonstrate. Features like the mean (μ), variance (μ_2), skewness (μ_3), kurtosis (μ_4), 5th moment (μ_5), 6th moment (μ_6), 7th moment (μ_7), standard deviation (σ), reciprocal of standard deviation ($1/\sigma$), root mean square (RMS), variance (σ^2), root sum of squares (RSS), sum of squares (SS), shape factor (w), impulsion index (I), crest factor (P), tolerance index (T), power to average (PA) are extracted from the raw data [2]. The definition of all these features is given in Table 5. In the table, x_ω represents the amplitude of frequency spectrum at a frequency of ω and N stands for the Nyquist frequency. Here, it must be noted that all the features mentioned (except $1/\sigma$) have standard definitions. However, $1/\sigma$ that has been used in this study is less common and often referred as precision. This feature retains the trends of standard deviation, but reduces its variation. Higher the spread, smaller the value of $1/\sigma$ would be. To demonstrate the same, Fig. 10 is plotted between variation of σ and $1/\sigma$ with each dataset for HP at 30 Hz speed. From the figure, it can be seen that there is a significant decrease in the variation of the feature.

Every fault in the CP contributes to changes in its flow patterns and thus has a unique effect on the CP signatures produced. In a multi-fault classification, it is critical to find feature(s) that is sensitive to every fault. Wrapper model is a performance-based feature selection technique. Therefore, to select the best features, a multi-class classification case of 15 faults is created, using HP, SB2, SB5, DB1, DB5, IFSB0, IFSB2, IFSB5, IFDB1, IFDB5, PCSB0, PCSB2, PCSB5, PCDB1, PCDB5 faults. These faults represent each fault family (as mentioned in Sect. 2) at intermediate severity (if applicable). The performance of different features in the aforementioned fault classification

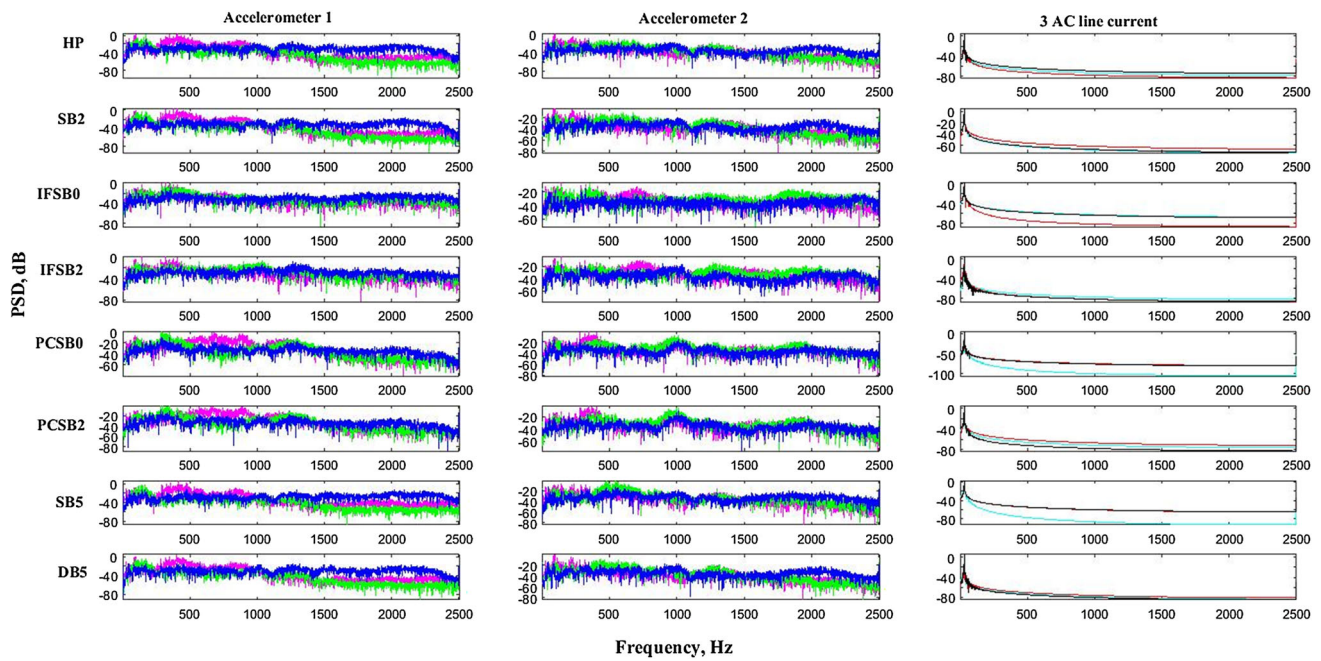


Fig. 9 Raw data from the accelerometers and the current probes of HP, SB2, IFSB0, IFSB2, PCSB0, PCSB2, SB5 and DB5 faults. In the accelerometer plots, magenta color represents radial transverse

direction, green color represents axial direction and blue color represents vertical transverse direction, whereas cyan, red and black are the three phases of motor line current

Table 5 Definition of various statistical features extracted from the frequency-domain data

$$\mu = \frac{1}{N} \sum_{\omega=0}^N x_{\omega}$$

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{\omega=0}^N |x_{\omega}|^2}$$

$$I = \frac{\max|x_{\omega}|}{\mu}$$

$$\mu_k = \frac{1}{N} \sum_{\omega=0}^N (x_{\omega} - \mu)^k$$

$$SS = \sum_{\omega=0}^N |x_{\omega}|^2 = \text{RSS}^2$$

$$P = \frac{\max|x_{\omega}|}{\text{RMS}} = \sqrt{PA}$$

$$\sigma = \sqrt{\mu_2}; \frac{1}{\sigma} = \frac{1}{\sqrt{\mu_2}}$$

$$w = \frac{\text{RMS}}{\mu}$$

$$T = \frac{\max|x_{\omega}|}{\left(\frac{1}{N} \sum_{\omega=0}^N |x_{\omega}|^2\right)^{1/2}}$$

at 30 Hz operating speed is shown in Fig. 11. From Fig. 11, it can be seen that μ , σ , $1/\sigma$, μ_3 , RMS and ω give a prediction performance of more than 85%. The combinations of these good performing features are then taken in triplets, and it is found that combination of μ , σ and $1/\sigma$ gives the best classification of 99.2%. Upon including an additional feature to this set of features (μ , σ and $1/\sigma$) also there is no improvement in the classification performance. This means that the combination of μ , σ and $1/\sigma$ features is able to precisely distinguish each fault type. Henceforth, a combination of these three features (μ , σ and $1/\sigma$) would be used for all the multi-fault classification cases. This feature selection has been done on the data acquired at the high-frequency resolution. However, to check the effect of data sampling resolution on the performance of the classifier, the same aforementioned faults along with the finalized

features are used for the comparison of performance of classifier using data sampled at both low- and high-frequency resolutions. The results of classification are given in Fig. 12. It can be clearly seen that the classification accuracy has improved significantly (by 8%) by using data from the high-frequency-resolution sampling. This may be attributed to the finer details captured by high-resolution data and thus better adaptation of the statistical features. Furthermore, owing to the better performance of the higher-resolution sampling, in the rest of the paper this would be used for various multi-fault classifications. It would be interesting to observe the scatter plot of various faults using these three features to understand their effectiveness in segregating the faults. For this purpose, a scatter plot of HP, SB2, SB5, DB1, IFSB0, IFSB2, PCSB0, PCSB2 faults at 45 Hz is shown in Fig. 13. It can be

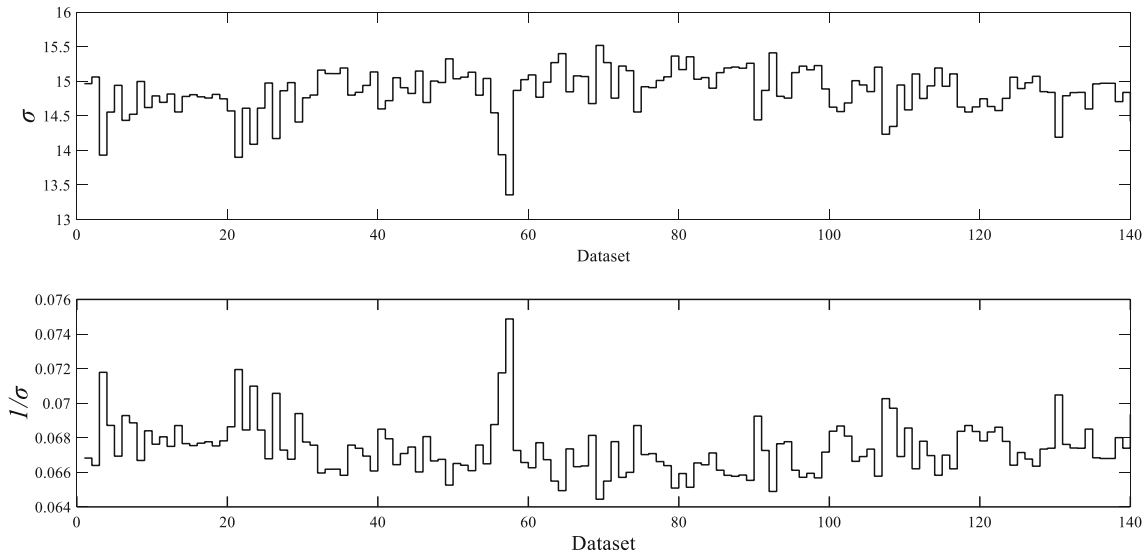


Fig. 10 Feature values of σ (above) and $1/\sigma$ (below) for each dataset of HP at 30 Hz

Fig. 11 Average classification performance of various features to segregate HP, SB2, SB5, DB1, DB5, IFSB0, IFSB2, IFSB5, IFDB1, IFDB5, PCSB0, PCSB2, PCSB5, PCDB1 and PCDB5 faults

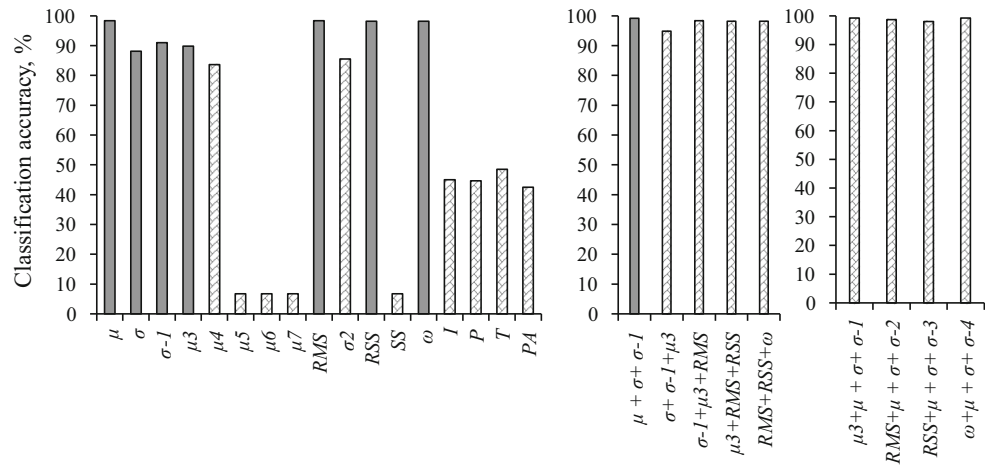
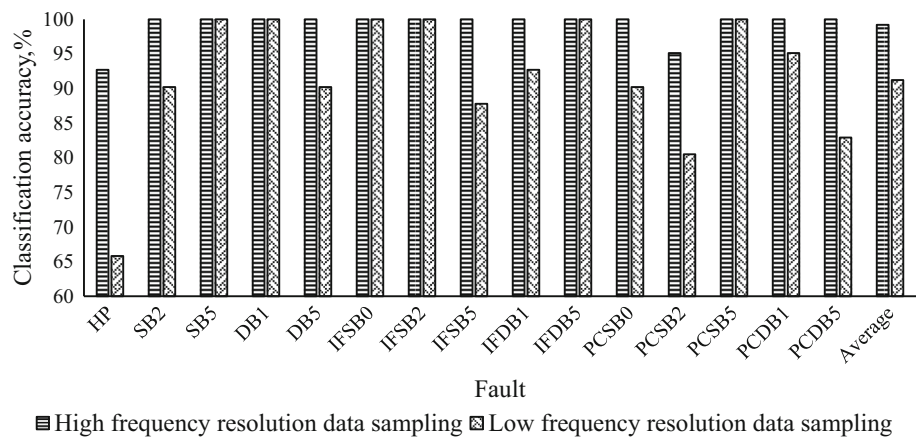


Fig. 12 Performance comparison of classifier using data sampled at high-frequency resolution and low-frequency resolution to segregate HP, SB2, SB5, DB1, DB5, IFSB0, IFSB2, IFSB5, IFDB1, IFDB5, PCSB0, PCSB2, PCSB5, PCDB1 and PCDB5 faults



observed from the figure that all the faults are very well segregated in separate clusters. There is a slight overlap of

features from HP, SB2, PCSB0 and PCSB2. This may be because of the closeness of severity levels of these faults.

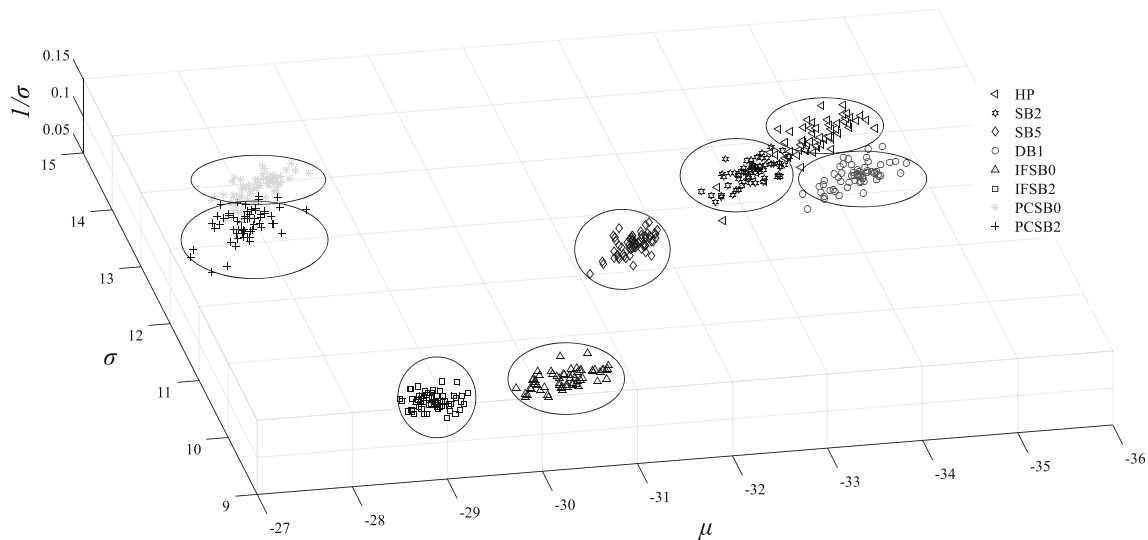


Fig. 13 Scatter plot of HP, SB2, SB5, DB1, IFSB0, IFSB2, PCSB0, PCSB2 faults at 45 Hz using features μ , σ and $1/\sigma$

4.1 Data required for training and testing

An estimate of the amount of data sufficient to train an MSVM model is critical. If a lot of data is used for training and less of it is used for testing, it leads to an ‘over-fit’ condition. If too few data are used for training and huge data are used for testing, it leads to an ‘under-fit’ condition. Hence, good number of samples needs to be taken each for the training and the testing of the MSVM algorithm. In the present investigation, the training/testing data ratios of—80/20, 70/30, 60/40, 50/50, 40/60, 30/70, 20/80—are taken. Here again, the fault classification of HP, SB2, SB5, DB1, DB5, IFSB0, IFSB2, IFSB5, IFDB1, IFDB5, PCSB0, PCSB2, PCSB5, PCDB1, PCDB5 faults is attempted at 30 Hz speed. It should be noted that the same number of samples is divided into different ratios in each case. μ , σ and $1/\sigma$ have been taken as a feature to perform the classification. The results are shown in Fig. 14. It can be seen

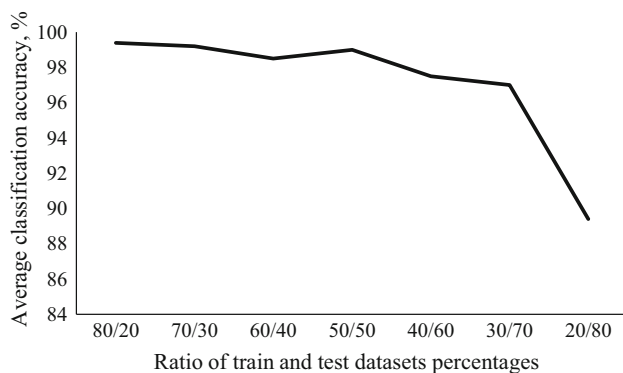


Fig. 14 Performance comparison of classifier to segregate HP, SB2, SB5, DB1, DB5, IFSB0, IFSB2, IFSB5, IFDB1, IFDB5, PCSB0, PCSB2, PCSB5, PCDB1 and PCDB5 faults with different train and test data ratios

from the figure that, as expected, there is a decrease in the fault classification performance with the decrease in the number of training samples. In addition, the performance of the classifier stabilizes at 60/40 and 50/50 sample ratios. Further, a decrease in training datasets is decreasing the classifier’s performance drastically. Also, if very less test samples are taken, the generalization of the algorithm decreases. Hence, for better generalization a 50/50 train test sample ratio is chosen for further analysis in this paper.

5 Frequency-domain prediction performance: results and discussion

In this section, results of the multi-class classification of various faults are presented. First different categories of fault classifications are given. Later, the significance of the considered fault classification category is illustrated and their results of classification are presented.

5.1 Fault classification cases and results

The multi-fault classifications attempted in this paper are: (a) pure CP blockage fault (BF) family; (b) CP impeller defect with blockage faults (IFBF) family, constituting IFSB0, IFSB1, IFSB2, IFSB3, IFSB4, IFDB1, IFDB2, IFDB3, IFDB4 and IFDB5; (c) CP with pitted cover plate and blockage faults (PCBF) family, constituting PCSB0, PCSB1, PCSB2, PCSB3, PCSB4, PCDB1, PCDB2, PCDB3, PCDB4 and PCDB5; (d) dry-run fault family, constituting SB5, IFSB5 and PCSB5 for different configurations of the CP; (e) all-fault (AF) classification case considers HP, SB2, SB5, DB1, DB5, IFSB0, IFSB2, IFSB5, IFDB1, IFDB5, PCSB0, PCSB2, PCSB5, PCDB1

and PCDB5 faults. Here, it is to be noted that AF case has been considered for the feature selection as well.

(a) BF fault classification

Blockage faults primarily constitute the suction and discharge blockages with varying severity levels. Both these faults have very diverse causes of formation. In addition, their existence results in discharge re-circulation and suction re-circulation (as explained in Sect. 1) in the CP. These faults are cumulative in nature and create cavitation-like damage on the CP.

In this case, an attempt is made to segregate BFs and identify their severities. The accurate recognition of these faults at their commencement will help take corrective measures to stop their progression. In this classification, ten faults are considered, four severities of suction blockages, including SB1, SB2, SB3 and SB4, five severities of discharge blockages, including DB1, DB2, DB3, DB4, DB5, and a healthy CP condition, HP. All these blockages are given on a healthy CP one at a time. The training and testing of the algorithm are done at the same operating speed of the CP. The results of this classification are presented in Table 6.

From the table, it can be seen that the average classification accuracy at all the speeds is above 91%. The overall classification accuracy is more than 96%. The best classification performance of 98.3% is found at 45 Hz speed. It can also be observed that the average classification accuracy is improving with the speed up to a speed of 45 Hz, after which there is no clear trend. This may be attributed to the stabilization of faults at higher speeds even though they are severe. The DB1 fault shows a perfect classification at all the speeds. The classification results are encouraging, and this algorithm seems effective in identifying each blockage fault type and segregates the discharge re-circulation and the suction re-circulation along with their severities.

(b) IFBF fault classification

Impeller cracks are mechanical faults; they not only create the CP unbalance but also disturb the fluid-flow pattern causing a pseudo-flow re-circulation (as explained in Sect. 2). They even act as stress-hot spots. If impeller defect is present in combination with fluid-flow blockages, it may result in an accelerated CP failure. In this classification case, eleven faults are considered, including (1) the impeller defect alone, (2) four levels of suction blockages on the CP with impeller defects, (3) five levels of discharge blockages on the CP with impeller defects and (4) the healthy pump condition. The reason for considering the HP condition is to check whether the algorithm is precisely able to distinguish between the faulty and non-faulty classes. It may be noted that the algorithm here is attempting to classify healthy pump (HP), mechanical faults (IFSB0 at low speeds), mechanical faults and pseudo-flow re-circulation (IFSB0, at higher speeds of CP operation), mechanical faults, pseudo-flow re-circulation combined with discharge re-circulation (IFSB1, IFSB2, IFSB3 and IFSB4), mechanical faults, pseudo-flow re-circulation combined with suction re-circulation (IFDB1, IFDB2, IFDB3, IFDB4 and IFDB5).

The fault training and testing are done at the same operating speed of the CP. The results of the classification are presented in Table 7. It can be seen from the results that the average classification accuracy at all speeds is above 92%. The overall classification accuracy is found to be 97.1%. The best classification performance of 99.6% is found at 60 Hz of CP speed. HP shows a perfect classification at all the speeds. IFSB0, IFSB2 and IFSB3 faults also show near-perfect classifications. It can also be observed that there is no clear trend in the fault classification performance, meaning that the algorithm is independent of the operating speeds of the CP. The complex fluid structure interactions may be the reason for some

Table 6 BF fault classification performance

Train Speed (Hz)	Test speed (Hz)	Fault classification accuracy										CV accuracy	Avg. accuracy	$\log_3 C$	$\log_3 \gamma$
		HP	SB1	SB2	SB3	SB4	DB1	DB2	DB3	DB4	DB5				
30	30	92.7	90.2	100	92.7	100	100	100	95.1	60.9	80.5	92.1	91.2	11	- 11
35	35	100	95.1	100	100	92.7	100	97.6	95.1	82.9	78	92.4	94.1	15.25	- 12.5
40	40	97.6	100	100	100	97.6	100	100	95.1	87.8	100	98.3	97.8	8	- 7.75
45	45	95.1	100	100	97.6	100	100	100	100	97.6	92.7	99.3	98.3	14	- 14
50	50	97.6	95.1	97.6	100	100	100	100	95.1	92.7	95.1	98.6	97.3	16.5	- 10.75
55	55	85.4	100	97.6	87.8	95.1	100	100	100	97.6	92.7	97.4	95.6	11	- 10
60	60	97.6	100	100	100	100	100	90.2	100	100	92.7	100	98	16.5	- 7
65	65	100	97.6	100	100	97.6	100	100	100	82.9	80.5	98.3	95.9	16.5	- 12

Table 7 IFBF fault classification performance

Train speed (Hz)	Test speed (Hz)	Fault classification accuracy										CV accuracy	Avg. accuracy	log ₃ C	log ₃ γ				
		HP	IFSB0	IFSB1	IFSB2	IFSB3	IFSB4	IFDB1	IFDB2	IFDB3	IFDB4					IFDB5			
30	30	100	100	100	100	100	100	100	100	100	100	100	100	100	100	96.9	97.3	15	- 14
35	35	100	100	97.6	100	100	97.6	100	100	100	100	100	100	100	100	99.1	99.3	16.5	- 11.75
40	40	100	100	100	97.6	100	90.2	92.7	97.6	100	97.6	100	100	100	100	96.1	95.4	15.5	- 12.75
45	45	100	97.6	100	100	97.6	100	97.6	100	60.9	65.8	97.6	100	100	100	98.3	92.5	15.5	- 14.25
50	50	100	100	100	100	100	100	100	100	95.1	48.8	97.6	87.8	100	100	95	93.6	13	- 12
55	55	100	100	100	100	100	100	100	100	97.6	95.1	100	100	100	100	97.4	98.9	12.5	- 11
60	60	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99.6	100	13	- 13
65	65	100	100	100	100	100	97.6	100	100	100	100	100	100	100	100	99.1	99.8	14	- 13.5

Table 8 PCBF fault classification performance

Train speed (Hz)	Test speed (Hz)	Fault classification accuracy															CV accuracy	Avg. accuracy	log ₃ C	log ₃ γ					
		HP	PCSB0	PCSB1	PCSB2	PCSB3	PCSB4	PCDB1	PCDB2	PCDB3	PCDB4	PCDB5													
30	30	100	100	100	87.8	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	96.8	91.1	13	- 11.25
35	35	100	95.1	80.5	97.6	97.6	100	97.6	100	100	100	97.6	90.2	82.9	85.4	95.1	95.1	95.7	92.3	8	- 8	95.7	92.3	8	- 8
40	40	100	100	95.1	100	100	100	100	100	100	100	95.1	97.6	85.4	95.1	87.4	96.9	96.9	95.8	13.25	13.25	96.9	95.8	13.25	- 12.5
45	45	100	100	100	100	100	100	100	100	100	100	92.7	95.1	90.2	100	73.2	97.4	97.4	95.6	12.5	12.5	97.4	95.6	12.5	- 11.5
50	50	100	92.7	100	80.5	100	100	68.3	100	100	68.3	100	100	85.4	85.4	100	94.2	94.2	92	12	12	94.2	92	12	- 10.75
55	55	100	97.6	97.6	97.6	100	100	95.1	100	100	95.1	100	100	100	100	97.6	98.9	98.9	98.7	14.5	14.5	98.9	98.7	14.5	- 14
60	60	100	92.7	97.6	100	100	100	85.4	100	100	85.4	29.7	87.8	92.7	100	92.2	92.2	89.6	89.6	15.5	15.5	92.2	89.6	15.5	- 13
65	65	100	97.6	100	100	97.6	100	97.6	100	100	97.6	100	100	90.2	95.1	75.6	97.2	97.2	95.8	11.5	11.5	97.2	95.8	11.5	- 10.25

Table 9 Dry-run fault classification performance

Train speed (Hz)	Test speed	Fault classification accuracy				CV accuracy	Avg. accuracy	$\log_3 C$	$\log_3 \gamma$
		HP	SB5	IFSB5	PCSB5				
30	30	100	100	100	100	100	100	16.5	- 6.75
35	35	97.6	100	100	100	100	99.4	16.5	- 4.25
40	40	100	100	100	100	100	100	16.5	- 4.5
45	45	100	100	100	100	100	100	16.5	- 5.75
50	50	97.6	100	100	100	100	99.4	16.5	- 5
55	55	100	100	100	100	100	100	16.5	- 5
60	60	100	100	100	100	100	100	16.5	- 3.5
65	65	100	100	100	100	100	100	16.5	- 5.25

faults not being able to achieve a 100% fault classification. However, the methodology seems to be very effective in identifying different flow instabilities and mechanical impeller faults.

(c) PCBF fault classification

Defects on the cover plate may result in turbulence inside the CP leading to a decrease in the pump performance. In addition, a combination of the pitted cover plate fault with blockage faults may lead to enhancement of the flow turbulence in the CP. In this case, the suction and discharge blockages are given individually on the CP with pitted cover plate faults. These faults are classified with respect to the HP condition and the PCSB0 condition. Like in the previous case of fault classification, in this analysis also mechanical faults and mechanical faults coupled with different types of re-circulation faults are classified. The training and the testing of the algorithm are done at the same operating speed of the CP. The results of the classification are presented in Table 8. From the table, it can be observed that at all speeds the classification accuracy is above 89.6%. In addition, the overall classification accuracy at all the speeds is found to be 93.9%. The highest classification of 98.7% is found at the CP operating speed of 55 Hz. The SB and PCSB4 faults give perfect classification at all the operating speeds. There is no clear trend of the fault classification, and hence, it can be said that the developed algorithm is versatile at all the operating speeds of the CP. The methodology is working desirably in classifying the mechanical faults as well as flow instabilities.

(d) Dry-run fault classification

In most CPs, priming is very essential. Non-priming of CP leads to a heat generation, which may result in bearing and seal failures. Hence, the dry run needs to be identified instantly. In this case, four faults are classified including dry run of a healthy CP (SB5), dry run of a CP with

impeller defects (IFSB5), dry run of CP with pitted cover plate faults (PCSB5) and a healthy pump condition (HP). The results of the classification are presented in Table 9. It can be observed from the table that the classification accuracy at all the speeds is above 99.4%. The overall classification accuracy for all the speeds is 99.9%. There is a perfect segregation of faults at speeds of 30, 40, 45, 55, 60 and 65 Hz. The IFSB5 and PCSB5 faults are perfectly classified at all operating speeds. There is no relation between the classification performance and the CP operating speed. Hence, the algorithm can be considered robust and can be used to readily identify the dry-run conditions.

(e) All-fault classification

All the classification cases presented so far considered segregation of faults within their specific family/group. Those cases would be applicable in places where CPs repeatedly fail due to a particular identified cause and it is important to recognize the fault severity. However, there may be cases where the causes of failure may be multiple and faults may exist simultaneously. In such cases, the present classification would be useful. In this classification case, fifteen faults, viz. no blockage, SB2, dry runs, DB1 and DB5, for all the three CP configurations along with HP condition are considered.

The algorithm is trained and tested at the same operating speed of the CP. Results of this classification are presented in Table 10. It can be seen from the table that the classification accuracy is above 96.9% at all the speeds. The overall classification accuracy at all the speeds is found to be 99.1%. DB1, DB5, IFSB2, IFSB5, PCSB0 and PCSB5 faults give a perfect classification at all the speeds. At speeds of 40 and 60 Hz, there is a 100% fault classification. There seems to be no correlation between the classification accuracy and the operating speed. It is interesting to note that this classification aims at identifying the pure mechanical faults, discharge re-circulation faults, suction

Table 10 AF fault classification performance

Train speed (Hz)	Test speed (Hz)	Fault classification accuracy															CV accuracy	Avg. accuracy	log ₃ C	log ₃ γ
		HP	SB2	SB5	DB1	DB5	IFSB0	IFSB2	IFSB5	IFDB1	IFDB5	PCSB0	PCSB2	PCSB5	PCDB1	PCDB5				
30	30	92.7	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99.8	99.2	16.5	- 7
35	35	100	100	100	100	100	100	100	97.6	100	100	100	100	97.6	100	100	99.5	99.5	15	- 16
40	40	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99.2	100	16.5	- 9.5
45	45	100	100	100	100	100	97.6	100	100	100	100	100	100	92.7	85.4	100	99.7	98.4	13	- 11.75
50	50	100	95.1	75.7	100	100	100	100	100	100	100	100	100	82.9	100	100	100	96.9	16.5	- 14.75
55	55	82.9	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99.7	98.9	16.5	- 10
60	60	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	16.5	- 7.5
65	65	100	100	100	100	100	100	100	100	100	100	100	100	95.1	100	100	99.8	99.7	16.5	- 8.25

Table 11 AF fault classification performance at intermediate speeds

Train speed (Hz)	Test Speed (Hz)	Fault classification accuracy															Avg. accuracy	C	γ				
		HP	SB2	SB5	DB1	DB5	IFSB0	IFSB2	IFSB5	IFDB1	IFDB5	PCSB0	PCSB2	PCSB5	PCDB1	PCDB5							
30, 40	35	88.7	23.4	100	100	100	100	100	100	100	100	100	100	99.3	69.5	90	100	100	95	99.3	91	1e7	5e-7
40, 50	45	92.2	87.2	82.9	100	87.9	90	100	100	100	100	100	100	91.5	61.7	100	0	0	0	0	86.2	1e7	5e-7
50, 60	55	73.2	92.7	100	100	31.7	65.8	100	100	95.1	100	100	100	29.3	0	0	0	0	0	0	72.5	1e7	5e-7

Table 12 Classification performance summary

Fault family	Train/test speed	Max. avg. accuracy/speed (Hz)	Min. avg. accuracy/speed (Hz)	Overall accuracy	Best classifiable fault(s)
BF	Same	98.3/45	92.1/30	96	DB1
IF	Same	99.6/60	95/50	97.1	HP, IFSB0, IFSB1, IFSB2, IFSB3
PC	Same	98.9/55	89.6/60	93.9	HP, PCSB4, PCSB5
Dry run	Same	100/30, 40, 45, 55, 65	99.4/35, 50	99.9	SB5, IFSB5, PCSB5
AF	Same	100/40,60	96.9/50	99.1	DB1, DB5, IFSB2, IFSB5, IFDB5, PCSB0, PCSB5
AF	Intermediate	91/(30-40), 35	72.5/50, 60	83.2	DB1, IFSB2, IFSB2

re-circulation faults, pseudo-re-circulation faults and mechanical faults combined with different aforementioned re-circulations at varied operating conditions of the CP. The performance of the methodology is very impressive.

Further, as a separate study the faults of this case are attempted to be classified at intermediate speeds, i.e., the algorithm is trained at two distinct speeds and tested at an intermediate speed. This is important in cases where the fault training data are not available at the present operating speed of the CP. For this, the algorithm is trained at 30–40, 40–50 and 50–60 Hz speeds, tested at 35, 45 and 55 Hz, respectively. The results of the classification are presented in Table 11. It can be seen from the table that the overall test classification accuracy is 83.2%. Clearly, the classification performance has decreased in comparison with the same speed training and testing condition. DB1, IFS2 and IFS5 faults are classified perfectly at intermediate speeds. At higher speeds, fault signatures get very distinct, and hence, the classification performance at higher intermediate speeds is decreasing. However, the results are promising, and the developed algorithm may be used for the fault identification at intermediate speeds.

5.2 Results summary

5.2.1 The summary of the results is tabulated in Table 12

To evaluate the competence of proposed methodology, results of the present work have been compared with other published literature. The comparison is tabulated in Table 13. The basis of comparison is sampling rate, signal domain, statistical features extracted, classifier used and its efficiency. From the present work, it is proven that the diagnosis of any possible independent or coexisting faults can be done using the present methodology with a high prediction rate. Also, the fault severity level may be estimated accurately. The prediction performance suggests that the established methodology can be applied to industrial problems.

6 Conclusions

A flexible algorithm has been proposed to classify the fault condition of the centrifugal pump based on multi-class support vector machine with hyperparameters optimization. The scope of this study is to find the best-suited features for the classification of thirty-three faults on the CP, including the suction and discharge blockage faults, impeller faults, pitted cover plate faults and dry-run faults. In this study, the interdependence of mechanical and hydraulic faults is considered. The hydraulic faults are considered at varying severities. A combination of centrifugal pump vibration data and the motor line-current data was used for the online monitoring of the pump condition. The wrapper model is used for the suitable statistical feature selection. The selected combined features of μ , σ and $1/\sigma$ are showing encouraging fault classification performance. Signature of every fault seeded on the CP was collected at two sampling rates of 20,000 S/s with 2000 S/dataset and 5000 S/s with 5000 S/dataset. Data sampled at high resolution (5000 S/s with 5000 S/dataset) gave better performance for fault classification, as it could capture the characteristic frequencies of the fault with better efficiency. However, if a higher resolution of data sampling is attempted to be used, then it must be kept in mind that both the sampling time and amount of storage space required for the data would increase substantially. Hence, 5000 S/s with 5000 samples per dataset seems to be enough for CP fault diagnosis. To check the robustness of the developed technique, the methodology is tested at eight different operating speeds of the CP. The developed algorithm could classify all the different types of flow instabilities, mechanical faults and the combination of both of these with very high precession at all the speeds. The reported classification performance of this study is much higher than that of the previous works done on combined mechanical and flow-induced fault diagnosis of CPs, making this methodology suitable for industrial applications. As a future work, an extensive study on the effect of

Table 13 Comparison of latest developments on CP fault diagnosis

Paper/year	Sampling rate	Types of faults	Domain used	Features considered	Classifiers/best prediction accuracy	Remarks
Rapur and Tiwari [5]	20 kHz	Suction blockages and impeller defects	Time-domain	Mean, peak to peak, skewness, kurtosis, standard deviation and entropy	SVM/99% (high blockage level)	1. Vibration signals used 2. Low suction blockage levels show low classification accuracies
Sakthivel et al. [8]	24 kHz	Bearing faults, seal defects, impeller defects, bearing and impeller fault together 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
Sakthivel et al. [9]	24 kHz	Bearing faults, seal defects, impeller defects, bearing and impeller fault together and cavitation	Time-domain	Mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range, minimum, maximum and sum	Decision tree/100%	1. Vibration signals used 2. Algorithm tested at only one operating speed of the CP
Wang [11]	10.24 kHz	Bearing roller wearing, bearing inner race wearing, bearing outer race wearing, and centrifugal pump impeller wearing	Time-domain	Sample entropy	CEEMD-SampEn and Random forests/97%	1. Vibration signals used 2. Algorithm tested at only one operating speed of the CP
Bordoloi and Tiwari [12]	20 kHz	Suction blockages with varying severity	Time-domain	Standard deviation, skewness, kurtosis	SVM with GA, ABCA, GSM optimization/94.7% (vs SVM with ABCA)	1. Vibration signals used 2. Classification accuracy of low suction blockages is 10%
Alfayez et al. [14]	10 Hz	Incipient cavitation	Time domain	N/A	N/A	Identified incipient cavitation using acoustic emission
Perovic et al. [15]	N/A	Cavitation, blockage, impeller defects with blockage, impeller defects with cavitation	Frequency domain	Slip, current amplitude and accumulative noise	Fuzzy system/83.7%	Motor current spectra was used
Rapur and Tiwari (current study)	20 kHz and 5 kHz	Suction blockage (with varying severity), Discharge blockage (with varying severity), impeller defects, impeller defects together with suction blockages, impeller defects together with discharge blockages, pitted cover plate faults, pitted cover plate together with suction blockages, pitted cover plate together with discharge blockages, dry run	Frequency domain	Mean, variance, skewness, kurtosis, 5th moment, 6th moment, 7th moment, standard deviation, reciprocal of standard deviation, root mean square, Variance, root sum of squares, sum of squares, shape factor, impulsion index, crest factor, tolerance index, power to average	SVM/100%	1. Vibration signals and motor line current used 2. Combined mechanical and hydraulic faults considered

choice of various kernel functions on the fault classification accuracies obtained may be studied. Since the fault signatures are usually non-stationary in nature, as a future direction time–frequency analysis using wavelet packets can be adapted for fault classification to improve the performance of the classifier.

Acknowledgements The authors are indebted to the reviewers and the editor for their valuable suggestions. The authors would like to thank the infrastructure and financial support provided by Indian Institute of Technology Guwahati, for carrying out the 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/> [41].

References

- Sun Y et al (2006) Mechanical systems hazard estimation using condition monitoring. *Mech Syst Signal Process* 20(5):1189–1201
- Tiwari R (2017) Rotor systems: analysis and identification. CRC Press, Boca Raton
- Hennecke F (2000) Reliability of pumps in chemical industry. In: Proceedings of pump users international forum, Karlsruhe, Germany
- Hart RJ (2002) Pumps and their systems—a changing industry. In: Proceedings of the international pump users symposium
- Rapur JS, Tiwari R (2017) Experimental time-domain vibration-based fault diagnosis of centrifugal pumps using support vector machine. *ASCE ASME J Risk Uncertain Eng Syst Part B Mech Eng* 3(4):044501–044507
- Tabar MTS, Majidi CH, Poursharifi Z (2011) Investigation of recirculation effects on the formation of vapor bubbles in centrifugal pump blades. *World Acad Sci* 49:883–888
- Muralidharan V, Sugumaran V (2013) Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump. *Measurement* 46(1):353–359
- Sakthivel NR, Sugumaran V, Nair BB (2010) Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump. *Mech Syst Signal Process* 24(6):1887–1906
- Sakthivel NR, Sugumaran V, Babudevasenapati S (2010) Vibration based fault diagnosis of monoblock centrifugal pump using decision tree. *Expert Syst Appl* 37(6):4040–4049
- Wolfram A et al. (2001) Component-based multi-model approach for fault detection and diagnosis of a centrifugal pump. In: Proceedings of the 2001 American Control Conference (Cat. No. 01CH37148)
- Wang Y et al. (2016) Fault diagnosis for centrifugal pumps based on complementary ensemble empirical mode decomposition, sample entropy and random forest. In: 2016 12th world congress on intelligent control and automation (WCICA)
- Bordoloi DJ, Tiwari R (2017) Identification of suction flow blockages and casing cavitations in centrifugal pumps by optimal support vector machine techniques. *J Braz Soc Mech Sci Eng* 1–12
- Chudina M (2003) Noise as an indicator of cavitation in a centrifugal pump. *Acoust Phys* 49(4):463–474
- Alfayez L, Mba D, Dyson G (2005) The application of acoustic emission for detecting incipient cavitation and the best efficiency point of a 60 kW centrifugal pump: case study. *NDT E Int* 38(5):354–358
- Perovic S, Unsworth PJ, Higham EH (2001) Fuzzy logic system to detect pump faults from motor current spectra. In: Conference record of the 2001 IEEE industry applications conference. 36th IAS Annual Meeting (Cat. No. 01CH37248)
- Cempel C (1988) Vibroacoustical diagnostics of machinery: an outline. *Mech Syst Signal Process* 2(2):135–151
- Kallesoe CS, Cocquemot V, Izadi-Zamanabadi R (2006) Model based fault detection in a centrifugal pump application. *IEEE Trans Control Syst Technol* 14(2):204–215
- Kong F, Chen R (2004) A combined method for triplex pump fault diagnosis based on wavelet transform, fuzzy logic and neuro-networks. *Mech Syst Signal Process* 18(1):161–168
- Gangsar P, Tiwari R (2017) Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms. *Mech Syst Signal Process* 94:464–481
- He M, He D (2017) A deep learning based approach for bearing fault diagnosis. *IEEE Trans Ind Appl* PP(99):1–1
- Fei S-W, Zhang X-B (2009) Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Syst Appl* 36(8):11352–11357
- Azadeh A, Ebrahimipour V, Bavar P (2010) A fuzzy inference system for pump failure diagnosis to improve maintenance process: the case of a petrochemical industry. *Expert Syst Appl* 37(1):627–639
- Huo Z et al. (2017) Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures. *IEEE Access*. PP(99):1–1
- Niu X et al (2017) Investigation of ANN and SVM based on limited samples for performance and emissions prediction of a CRDI-assisted marine diesel engine. *Appl Therm Eng* 111:1353–1364
- Jiao X et al (2017) Research on fault diagnosis of airborne fuel pump based on EMD and probabilistic neural networks. *Microelectron Reliab* 75:296–308
- Zheng J, Pan H, Cheng J (2017) Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mech Syst Signal Process* 85:746–759
- Batista FB et al (2016) An empirical demodulation for electrical fault detection in induction motors. *IEEE Trans Instrum Meas* 65(3):559–569
- Rapur JS, Tiwari R (2017) A compliant algorithm to diagnose multiple centrifugal pump faults with corrupted vibration and current signatures in time-domain. In: ASME 2017 Gas Turbine India Conference. American Society of Mechanical Engineers, pp. V002T05A007–V002T05A007
- Zhao W et al. (2016) Fault diagnosis for centrifugal pumps using deep learning and softmax regression. In: 2016 12th world congress on intelligent control and automation (WCICA)
- Vladimir VN, Vapnik V (1995) The nature of statistical learning theory. Springer, Heidelberg
- Widodo A, Yang B-S (2007) Support vector machine in machine condition monitoring and fault diagnosis. *Mech Syst Signal Process* 21(6):2560–2574
- Behzad M, Bastami AR, Maassoumian M (2004) Fault diagnosis of a centrifugal pump by vibration analysis 41758:221–226
- Wang J, Hu H (2006) Vibration-based fault diagnosis of pump using fuzzy technique. *Measurement* 39(2):176–185
- Harihara PP, Parlos AG (2006) Sensorless detection of cavitation in centrifugal pumps. In: Proceedings of 2006 ASME international mechanical engineering congress and exposition. Chicago, USA: ASME

35. Harihara PP, Parlos AG (2008) Sensorless detection of impeller cracks in motor driven centrifugal pumps. In: ASME 2008 international mechanical engineering congress and exposition. American Society of Mechanical Engineers
36. Mba D, Rao RB (2006) Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines; bearings, pumps, gearboxes, engines and rotating structures. *Shock Vib Dig* 38:3–16
37. Venkatachalam R (2014) *Mechanical Vibrations*. PHI Learning Pvt Ltd, New Delhi
38. Wang H, Chen P (2007) Fault diagnosis of centrifugal pump using symptom parameters in frequency domain. *Agric Eng Int CIGR J* 9:1–14
39. Dular M, Stoffel B, Širok B (2006) Development of a cavitation erosion model. *Wear* 261(5):642–655
40. Azadeh A, Saberi M, Kazem A, Ebrahimipour V, Nourmohammadzadeh A, Saberi Z (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. *Appl Soft Comput* 13(3):1478–1485
41. Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2(3):1–27
42. Frazer, H (1981) Flow recirculation in centrifugal pumps. In: ASME meeting