



Cavitation Detection of Centrifugal Pumps Using SVM and Statistical Features

Berli Kamiel^(✉) , Taufiq Akbar , Sudarisman, and Krisdiyanto

Mechanical Engineering Department, Universitas Muhammadiyah Yogyakarta, 55183
Yogyakarta, Indonesia
berlikamiel@umy.ac.id

Abstract. Centrifugal pump is one type of pump that is widely used in the industry. Its mechanism which creates pressure change may cause cavitation. Generally, the cavitation increases noise and vibration level, if not properly maintained leads to catastrophic failure and total stop of the whole process. Therefore, it is needed to develop a method that can detect cavitation as early as possible. Recently, pattern recognition-based fault detection is gaining attention from many researchers due to its superior detection and classification performance. Support vector machine (SVM) is one of the pattern recognition techniques which requires statistical features as input for building the model. However, the selection of statistical features is arbitrary. In this study, ten statistical features are extracted from the time-domain vibration signal and selected using Relief Feature Selection. The selected features are used as input for two types of SVM, binary, and multi-class, to classify new vibration data. Each multi-class classification result is optimized by the Bayesian Optimization algorithm and Grid Search Method. The result shows that root mean square, standard deviation, variance, entropy, and standard error are several features that indicate the best plot. The feature selection process reveals that variance, root mean square, and standard error are the best feature to use for SVM classification. The binary SVM method shows the best plot on early cavitation with an accuracy of 99%. The Bayesian optimization algorithm with multi-class SVM is the best combination to classify all pump conditions with an overall accuracy of 100%.

Keywords: Centrifugal pump · Support vector machine · Relief feature selection · Bayesian optimization · Grid search method

1 Introduction

Centrifugal pump is one type of pump widely used in the industry such as chemical, oil, and gas company because of its simple mechanism and construction. Considering the importance of its role, the performance is salient to be maintained. The reduction of its performance is usually affected by faulty components. One of the most common causes of components' failure is cavitation. Cavitation occurs due to the evaporation of liquid flowing fluid which is caused by a pressure drop below the saturated vapour pressure.

The cavitation diagnosis is very important to be carried out at an early stage because a decrease in pump capacity due to cavitation can disrupt production activities. This will certainly impact the level of productivity in an industry so that an effective method is needed in detecting early cavitation in centrifugal pumps.

Early cavitation detection on the centrifugal pump was examined and tested by several researchers in various methods. Al-Hashmi et al. [1] utilized spectrum analysis from vibration signals. On the other hand, Al-Obaidi [2] examined the use of statistical features in the time domain to detect faults in centrifugal pumps. Another approach was introduced by Nasiri et al. [3] which uses Artificial Neural Network (ANN) based vibration signal for detecting cavitation in a centrifugal pump. Sakthivel et al. [4] applied the Decision Tree algorithm in detecting cavitation in a centrifugal pump. Farokhzad [5] detected several faults of the pump using the Adaptive Network Fuzzy Inference System (ANFIS); however, the indication of cavitation at the initial stage was difficult to observe.

The researchers then try to use one of the pattern recognition-based methods, called support vector machine (SVM). This method is considered as the method which enables to classify data and provide information with a high level of accuracy. Ebrahimi and Javidan [6] investigated faults in centrifugal pumps using vibration signal analysis and the SVM method. Vibration signals were decomposed in three-level using wavelet transform, and descriptive statistical features were extracted from detail and approximation coefficients of the wavelet. Saberi et al. [7] found that the SVM method had the advantage of having a kernel function that could distinguish normal conditions and high noise on the centrifugal pump. Sakthivel et al. [8] compared the use of Proximal Support Vector Machine (PSVM), Gene Expression Programming (GEP), Wavelet-GEP, and SVM, which showed that SVM was the method with the highest accuracy.

The use of SVM in detecting early cavitation requires the help of a kernel function. The selection of the right kernel function affects the detection results. The recommended kernel function is Gaussian Radial Basis Function (RBF) because it can classify non-linear data groups [9]. In addition, the statistical features in the time domain used can also affect SVM performance. Rapur and Tiwari [10] proposed that the statistical parameters mean, standard deviation, and entropy because they give excellent accuracy. On the other hand, Elangovan et al. [11] proved that the promising statistical parameter is the standard error and minimum value.

However, there is no standard procedure in determining the use of statistical parameters, and the best classification algorithm for detecting cavitation in a centrifugal pump is the SVM. In this study statistical parameters are extracted from the time domain instead of the frequency domain since the spectrum from the vibration of fluid interaction in centrifugal pumps is dominated by noise and random high-frequency vibration which has less meaningful information. Further research and development of the SVM-based method are needed to get better accuracy. Hence this paper aims to find the best combination of the statistical features, and optimization algorithm in SVM to detect early cavitation and to classify several levels of cavitation in the centrifugal pump.

2 Support Vector Machine

SVM is one method that classifies data based on pattern recognition. Pattern recognition works with separating data into several groups or classes [12]. This method can be classified as part of an artificial intelligence system built for decision-making. Input data is used in various, such as numbers, images, sounds, or a signal wave. This method is very popular in the field of statistics. Up to now, the implementation of the analysis vibration signal method based on pattern recognition continues to develop. This affects the increasing number of new methods based on pattern recognition. The appearance of this new method indicates that the level of popularity in the future will be even better.

SVM is a method that is used for binary classification. Originally, SVM combined several sets of concepts in the field of pattern recognition. This method works by finding the best hyperplane that separates groups of data on a dimension perfectly into two classes. The pattern of the two classes assumed has been completely separated by a hyperplane in a dimension defined by Eq. (1).

$$\vec{w} \cdot \vec{x} + b = 0 \quad (1)$$

If \vec{x}_i is in class -1 , then as shown in Eq. (2).

$$\vec{w} \cdot \vec{x} + b \leq -1 \quad (2)$$

While the value \vec{x}_i in class $+1$ is shown in Eq. (3).

$$\vec{w} \cdot \vec{x}_i + b \geq +1 \quad (3)$$

Quadratic Programming (QP) problem is an effort to maximize the value of the distance between the hyperplane and its closest point by finding its minimum point like Eq. (4).

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} w^2 \quad (4)$$

$$y_i(\vec{x}_i \cdot w + b) - 1 \geq 0, \forall_i \quad (5)$$

The problem in Eqs. (4) and (5) can be solved by the Lagrange Multiplier (α_i) technique, as in the following Eq. (6).

$$L(\vec{w}, b, \alpha) = \frac{1}{2} \vec{w}^2 - \sum_{i=1}^l \alpha_i (y_i (\vec{x}_i \cdot \vec{w} + b) - 1); (i = 1, 2, \dots, l) \quad (6)$$

Furthermore, Eq. (4) is optimized by maximizing problem that only contained value (α_i), as in the following Eq. (7).

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j x_i x_j \quad (7)$$

So, Eq. (8) is obtained:

$$\alpha_i \geq 0 (i = 1, 2, \dots, l) \sum_{i=1}^l \alpha_i y_i = 0 \quad (8)$$

Equation (8) shows that the value (α_i) is mostly positive, and the support vector correlates with this positive value (α_i).

Figure 1 shows several patterns of two classes (the pattern in class A is marked with a red box symbol and the pattern in class B is marked with a blue circle symbol) which will further be processed to search for the best hyperplane by giving rise to several discrimination boundaries (alternative split lines). In determining the best hyperplane, it can be done by finding the maximum point and measuring the margin. Figure 2 shows the best hyperplane located in the middle between the two classes, and the pattern near the hyperplane is Support Vector.

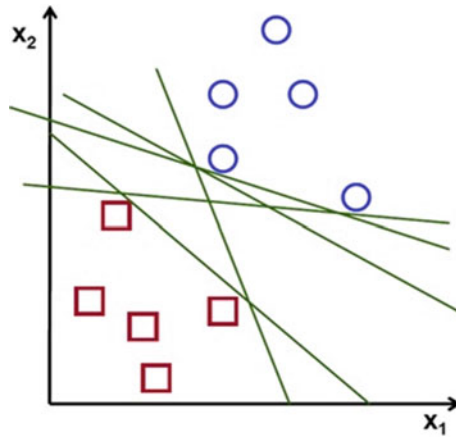


Fig. 1. Search for optimal hyperplane

3 Method

The data acquisition process was taken by recording 500 files of data from vibration signal for every four variations of pump conditions, including normal (0% valve blockage), level 1 cavitation (0.25% valve blockage), level 2 cavitation (0.50% valve blockage), and level 3 cavitation (0.75% valve blockage). The recording time was 10 s per file and paused for 2 s between recordings. The pump rotational speed was regulated at 2850 RPM, and a sampling rate of 17,066 Hz was set to produce a stable change in each condition. The recording was carried out using the NI 9234 acquisition device with the chassis NI DAQ 9174, and an accelerometer mounted on the pump inlet. The data acquisition process was regulated using the NI MAX and Matlab R2017a. The test rig setting consists of main components such as a centrifugal pump, closed-loop pipe network, gauges valves, water tank, and flowmeter.

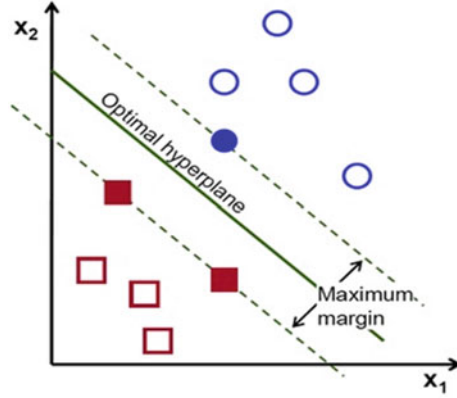


Fig. 2. Optimal separating hyperplane

3.1 Feature Extraction and Selection Process

The vibration signal was extracted into ten statistical features in the time domain as shown in Table 1. Each statistical feature was then plotted to show its characteristics to the distribution of data from the vibration signal of all conditions. All results of the extraction process were then selected. The feature selection process was done by using Relief Feature Selection. This method ranked the statistical features based on the weight of the information content. The feature selection process then produced the best data input for SVM classification.

Table 1. Statistical features in time domain

Statistical features	Formulas	Statistical features	Formulas
Root Mean Square (RMS)	$\sqrt{\frac{1}{n} \sum_{i=1}^N (x_i - \bar{x})^2}$	Crest factor	$\frac{\max(x(n))}{\sigma}$
Standard Deviation (SD)	$\sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$	Mean	$\frac{1}{K} \sum_{k=1}^K x(k)$
Peak value	$\max x(N) $	Entropy	$\sum_{i=1}^N p(x_i) \log_{10} p(x_i)$
Kurtosis	$\frac{\sum_{i=1}^N (x_i - \bar{x})^4}{(N-1)\sigma^4}$	Minimum value	$\min x(N) $
Variance	$\frac{(x_i - \bar{x})^2}{N-1}$	Standard Error (SE)	$\sqrt{\frac{1}{(n-2)} - \left[\sum (y - \bar{y})^2 - \frac{[\sum (x - \bar{x})(y - \bar{y})]^2}{(x - \bar{x})^2} \right]}$

3.2 Classification Process

The classification stage was carried out using the SVM-based method. To determine the various levels of cavitation and the initial formation, two methods were applied, namely Binary and Multi-Class SVM. Kernel Function for mapping process used Radial Basis Function (RBF). The multi-class SVM classification method was carried out with three trials with an optimization method. The optimization method used at this stage was Grid Search Method (GSM) and Bayesian Optimization (BO) techniques.

GSM algorithm is an optimization technique based on grid search. Grid search is performed on each mapping function. Each mapping function with optimal results and not optimal will be evaluated by the GSM algorithm so that the number of evaluations produced in several mapping processes that occur. The results of optimization then are sorted according to functions that have the best classification parameters.

BO techniques were carried out to evaluate any mapping errors found. This technique was equipped with an acquisition function that was useful for determining which mapping function was not optimal. Then the mapping function was optimized and would be returned in the training process. The number of mapping functions evaluated according to the number of mapping error processes that occurred.

4 Result

Statistical features in time domain aim to find characteristics of data from vibration signals under normal conditions, level 1, 2, and 3 cavitation. Some features such as RMS, SD, variance, entropy, and SE show the separation of the pump conditions. However, for normal conditions and level 1 cavitation, they fail to separate the classes. Thus, the selection process is intended to select the best features as input for SVM classification.

The result of features selection using Relief Feature Selection can be seen in Fig. 3. The result of the selection shows that the variance has the highest weight meanwhile the crest factor is the lowest one. Three features with the highest scores were used as SVM input since they contained a weight of more than 0.0050. However, the SE is not used as input because it has the same value as SD.

The binary SVM classification was performed on four pump conditions. The two main stages carried out in this classification are the training and testing process. The training forms a classification model and generates a mapping function, while testing data evaluates the training model and determines the accuracy. The input data was separated using the cross-validation process to avoid overfitting. This process includes 800 samples for training and 200 samples for the testing process.

The steps taken in the multi-class SVM correspond to the binary SVM method. The method classifies the four pump conditions and shows the result of the spread of the pattern. The cross-validation process produces 1800 training data and 200 testing data, where all variations have the same amount of data in each set.

The classification was done by three experiments, those were multi-class SVM without optimization, using GSM optimization, and BO algorithm. From the experiments, the highest level of accuracy is shown by a combination of multi-class SVM and BO.

Based on the distribution of statistical parameters, RMS, SD, variance, entropy, and SE solely can separate four pump conditions properly with an exception for normal and

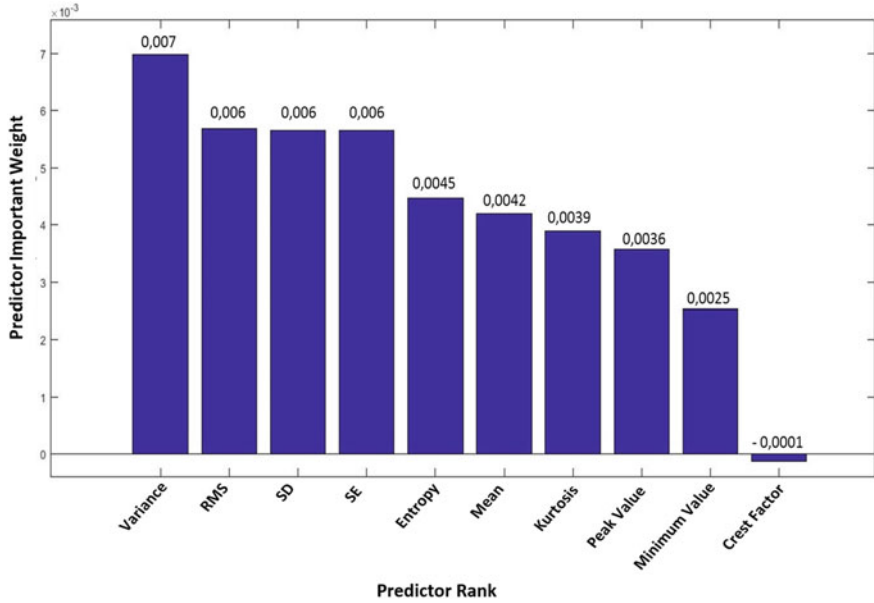


Fig. 3. Features selection result

level 1 cavitation classes which show less clustering effect. However, the results have a promising potential to be used as the input parameters for SVM.

Table 2 shows the success of SVM binary classification in detecting early cavitation phenomena, with an accuracy of 99%. The separation of the four classes is carried out using a combination of multi-class SVM and BO as depicted in Table 3 with a maximum accuracy of 100%. The obtained classification accuracy is a bit higher compared to that from [5] where they extracted statistical parameters from wavelet coefficients. The relief feature selection procedure is proven to be more effective to choose suitable statistical parameters than the wavelet-based parameters extraction.

Table 2. Binary SVM accuracy

No	Binary SVM	Accuracy (%)
1	Normal and Level 1 cavitation	99
2	Normal dan Level 2 cavitation	97
3	Normal dan Level 3 cavitation	100

The results of the study reveal that the SVM-based cavitation detection method show the early cavitation phenomenon. The combination of the SVM-BO algorithm is the most superior method to detect cavitation in several levels.

Table 3. Multi-Class SVM accuracy

Method	Pump condition	Misclassification	Accuracy (%)	Overall accuracy (%)
Multi-class SVM (without optimization)	Normal	23	13.50	58
	Level 1	13	18.50	
	Level 2	15	17.50	
	Level 3	33	8.50	
Multi-class SVM (GSM optimization)	Normal	0	25	99.50
	Level 1	0	25	
	Level 2	1	24.50	
	Level 3	0	25	
Multi-class SVM (Bayesian optimization)	Normal	0	25	100
	Level 1	0	25	
	Level 2	0	25	
	Level 3	0	25	

5 Conclusion

The study indicates that the character of each statistical feature in the time domain produces specific information to the vibration signal distribution. RMS, SD, and variance are the most suitable statistical features used as input for SVM classification. The SVM has been proven to detect early cavitation phenomena in the centrifugal pump. This is shown in the classification between normal and level 1 cavitation that has an accuracy of 99%. Development and optimization of SVM multi-class algorithms by using BO is the best combination method in detecting cavitation at several levels. The level of accuracy obtained with this combination is 100%.

References

1. Al-Hashmi S, Abdurrrhman A, Nasir I, Elhaj M (2017) Spectrum analysis of vibration signals for cavitation monitoring. *J Pure Appl Sci. Sebha University*: 16 14–20
2. Al-Obaidi AR (2020) Detection of cavitation phenomenon within a centrifugal pump based on vibration analysis technique in both time and frequency domains. *Exp Tech* 44(3):329–347
3. Nasiri MR, Mahjoob MJ, Vahid-Alizadeh H (2011) Vibration signature analysis for detecting cavitation in centrifugal pumps using neural networks. *IEEE Int Conf Mechatron* 13–15
4. 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
5. Farokhzad S (2013) Vibration based fault detection of centrifugal pump by fast fourier transform and adaptive neuro-fuzzy inference system. *J Mech Eng Technol*: 82–87
6. Ebrahimi E, Javidan M (2017) Vibration-based classification of centrifugal pumps using support vector machine and discrete wavelet transform. *J Vibroengineering* 19(4): 2586–2597

7. Saberi M, Azadeh A, Nourmohammadzadeh A, Pazhoheshfar P (2011) Comparing performance and robustness of SVM and ANN for fault diagnosis in a centrifugal pump. In:19th International Congress on Modelling and Simulation, Perth
8. Sakthivel NR, Binoy.B.Nair, Sugumaran V (2012) Soft computing approach to fault diagnosis of centrifugal pump. *Appl Soft Comput* 12(5): 1574–1581
9. Sakthivel NR, Saravanamurugan S, Nair BB, Elangovan M, Sugumaran V (2016) Effect of kernel function in support vector machine for the fault diagnosis of pump. *J Eng Sci Technol* 11(6): 826–838
10. Rapur JS, Tiwari R (2017) Experimental time-domain vibration-based fault diagnosis of centrifugal pumps using support vector machine. *ASCE-ASME J Risk Uncert Eng Sys Part B Mech Eng* 3(4)
11. Elangovan M, Sugumaran V, Ramachandran KI, Ravikumar S (2011) Effect of SVM kernel functions on classification of vibration signals of a single point cutting tool. *Expert Syst Appl* 38(12):15202–15207
12. Mavroforakis ME, Theodoridis S (2006) A geometric approach to support vector machine (SVM) classification. *IEEE Trans Neural Netw*17(3): 671–682