APPLICATION OF SUPPORT VECTOR MACHINE FOR EVALUATION OF WEAR STATE AND REMAINING LIFE TIME

Mahmud-Sami Saadawia, Dirk Söffker

University of Duisburg-Essen, Chair of Dynamics and Control
Lotharstr.1-21, 47057, Duisburg, Germany
Email: mahmud-sami.saadawia@uni-due.de, soeffker@uni-due.de

Abstract

Usually, mechanical machines do not break down or fail without any kind of warning, which is indicated by a combination of changing measurable symptoms. The complexity and the high dimensionality of the measured signals require reliable, fast, and less demanding methods to recognize the faults. In this work, a production machine related supervision task is investigated over a long duration to design a fault detection and prediction system to support condition-based maintenance of wear parts and to detect and predict failures usually leading to the full loss of functionality. Wear parts failure should be detected before scuffing or seizing lead to serious failure of the machine. An approach for developing the system as a prewarning module is presented. The system is based on support vector machine (SVM) classification as a signal-based diagnosis technique and as a feature fusion tool. The processed and extracted parameters of the machine operation are investigated and fused by the SVM to find the most reliable features for the detection system. Alternative combinations of fusing sensors are taken into consideration to find a complementary sensor array for better accuracy. A parameter indicating the need for wear part replacement; a Change Index (CI) is presented based on the decision value resulting from the SVM which shows a tendency to change over time coinciding with the deterioration of the part and the remaining life time.

Keywords: Condition monitoring, SVM classification, Signal based modeling, Feature fusion, Fault detection and prediction.

Presenting Author's Biography

Mahmud-Sami Saadawia, was born in Benghazi, Libya, in 1969. He received the B.Sc. degree in Mechanical Engineering from the University of Garyounis, Benghazi in 1993, and the M.Sc. degree in Mechatronics from the University of Duisburg-Essen in 2006, and he is currently a Ph.D. student of the Chair of Dynamics and Control, University of Duisburg-Essen. His current research interests include SVM, wavelets, and applications of machine learning techniques in fault diagnosis and prognosis.

1 Introduction

The use of condition-based maintenance has been addressed using a variety of approaches in order to reduce the costs of industrial systems and to maximize the usage of machinery parts and systems. Condition-based maintenance indicators describing the condition of a machine or machine parts are used to determine the state of the system which can be used to optimize maintenance procedures, system operation, and life cycles.
A machine will fail with time due to the wear and ultimate failure of its critical components. The question is, when will this occur? Practically, an exact answer is very difficult even with availability of many measurements and operation parameters, but studying such measurements and operation parameters of existing and operating machines can give a reliable warning to avoid unexpected or severe problems.
Usually, mechanical machines do not break down or fail without any kind of warning, which is indicated, for example, by an increased vibration level, increased hydraulic pressure, decreased displacement, or a combination of all these phenomena. On the other hand, the complexity and the high dimensionality of the measured signals require reliable, efficient, fast, and less demanding methods, which can be easily validated, to recognize the faults by measuring symptoms. Many approaches have been used such as knowledge-based systems, model-based control systems, and statistical approaches. The use of signal-based approaches in machine learning techniques has been providing comparatively good performance without the need for complex modeling task necessary for model-based approaches. Recently, the concept of information fusion has been added to the above mentioned techniques and has been used to improve the accuracy of recognizing faults.
In this work, a production machine related supervision task is investigated to design a fault detection system based on SVM to detect failures usually leading to the full loss of functionality. An indication of the remaining life of the material is also required. Wear failure should be detected before scuffing or seizing lead to serious failure of the machine. The parameters of the machine operation are investigated and fused by Support Vector Machine (SVM) to find the most reliable features for the detection system. Alternative combinations of fusing sensors are taken into consideration to find a complementary sensor array for better accuracy.

2 Learning with SVM

After introduced by Cortes and Vapnik [1], based on statistical learning theory, SVM spread after the comparatively excellent results were achieved in text recognition and image classification [2, 3]. Since then SVM has become a very popular technique for classification and pattern recognition. Since early applications in fault diagnosis [4, 5], Support Vector Machine (SVM) has been showing better results compared to other capable techniques such as neural networks and model-based reasoning.

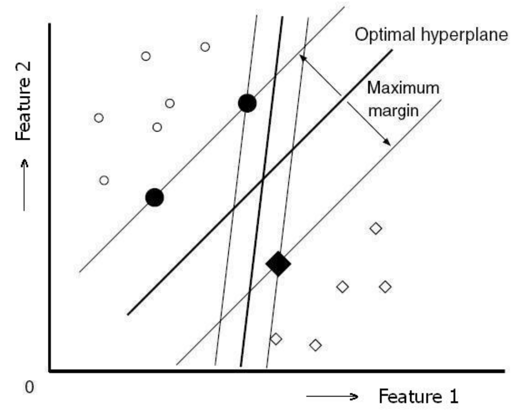

Fig. 1 The maximum margin separating hyperplane [6]

2.1 Mathematical description

The learning problem setting of SVM [6] is to find the unknown nonlinear dependency (mapping, function) between some high dimensional input vector x and scalar output y or as vector output y used in multiclass SVM. In general there is no information about the underlying joint probability function [7]. Thus one must perform a distribution-free learning. The only information available is a training data set. In general [6] SVM involves and depends on the solution of the quadratic optimization problem
to minimize

$$J = \frac{1}{2}W^T W + C\sum_{i=1}^{l} \xi i \quad (1)$$

with respect to

$$y_i(W^T \phi(x_i) + b) \geq 1 - \xi_i \quad (2)$$

and

$$\xi_i \geq 0, \quad (3)$$

where W represents the coefficient vector of the separating hyperplane, C represents a penalty parameter, ξ_i is a slack variable associated with the data point x_i where the number of the data points is l, b represents the bias term of the separating hyperplane, and ϕ is a mapping function. A decision function $D(x)$ is used to classify the unknown data points according to the position and distance from the separating hyperplane. The decision function value is coinciding with the distance from the separating hyperplane and used as follows;
The unknown data point x is classified into

$$\begin{cases} Class1 & \text{if} \quad D(x) > 0 \\ Class2 & \text{if} \quad D(x) < 0 \end{cases} \quad (4)$$

If $D(x) = 0$, x is on the boundary and thus is unclassifiable. The region

$$x|1 > D(x) > -1 \quad (5)$$

is the generalization region of the classifier.
The function

$$K(x_i, y_j) = \phi(x_i)^T \phi(x_j) \quad (6)$$

is called the kernel function. Kernel functions are used to map the input data from the input space into a higher dimensional feature space, where the separating hyperplane is constructed. Here SVM finds a linear separating hyperplane with the maximal margin (Fig. 1).

2.2 Advantages of SVM

The most important advantage of the SVM is the generalization ability. This is because of the maximum margin criterion in the process of selecting the separating hyperplane. Support Vector Machine is trained to maximize the margin, thus the generalization ability is better under conditions such as scarce training data. Additionally, the feature space approach of the SVM can be a tool to realize a complementary sensor array for feature fusion [8], where a combination of signals provides a more complete information of the problem and therefore better accuracy than individual signals.
Another advantage of SVM is its robustness to outliers. Proper setting of the penalty parameter C which controls the misclassification error suppresses the outliers and reduces the effect of increasing noise level. In neural networks, for comparison, the outliers need to be eliminated before training [6].

2.3 Multisensor data fusion

Multisensor data fusion [8] refers to the intelligent processing of a set of two or more sensors that have different levels of cooperative, complementary, competitive, and independent qualities. Cooperative sensors are combining sensors, which gives a new quality of information while complementary sensors are those fused to give more clear information of the problem. A competitive array provides unrelated measurements of the same physical phenomena. Data fusion are classified in three levels: signals fusion, features fusion, and decision fusion. SVM can be used as a tool for features fusion level where extracted features are fused in the feature space to give better information than individual features.

3 The given input data

Vibration measurement is becoming increasingly popular as a condition-based monitoring procedure and as support for machinery maintenance decisions. Here the velocity of vibration is a very important item to measure medium frequencies (until about $1kHz$), where the failure mode is fatigue and wear out of surfaces. On the other hand, the velocity of vibration is sometimes not suitable in extracting transient, process-related, or impulse-like incidents, where the localized high frequencies are dominating for short times. The sensor data includes, in addition to vibration velocity and vibration acceleration, for the considered system two other important properties: the system hydraulic pressure and the displacement of the piston of the monitored parts. Combining both provides implicit information about the friction and hence the tribological condition of the parts surfaces.

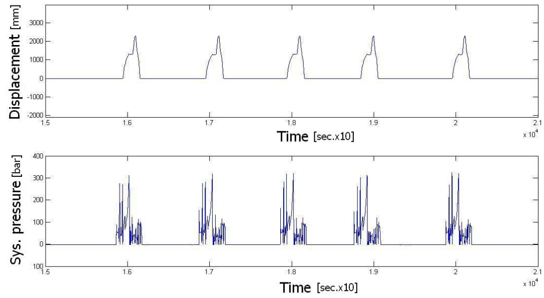

Fig. 2 Sample of raw data: displacement and system pressure

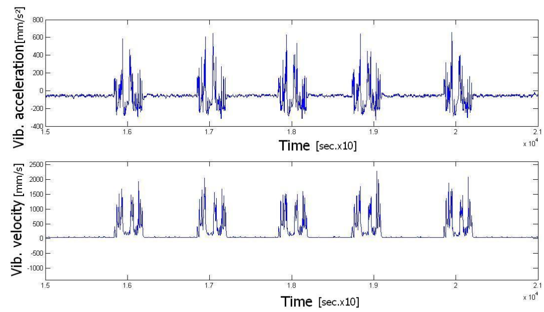

Fig. 3 Sample of raw data: vibration velocity and vibration acceleration

4 Data preparation and feature extraction

In order to apply a successful classification process, the data must be prepared by careful transformation and feature extraction procedures. The aim of such procedures is to exclude redundant information and to put the useful information for classification in an recognizable structure.

4.1 Feature extraction

The time series vectors of the sensory data (Fig. 2 and Fig. 3) do not comply with the cyclic nature of the considered machine. Indicators of classification can be parts of the physical cycle of the machine, for example the high starting vibration velocity as given in Fig. 4. In this case, if classification is applied to time series vectors, the other parts of the cycle will deteriorate the efficiency of the classification (Fig. 5). This is because no difference between these points according to machine states is observed. After cleaning off the redundant signals, the signals should be structured so that transformed observations are as indicative as possible of the system states in the classification. Extraction of the operation cycles of the machine is a preliminary

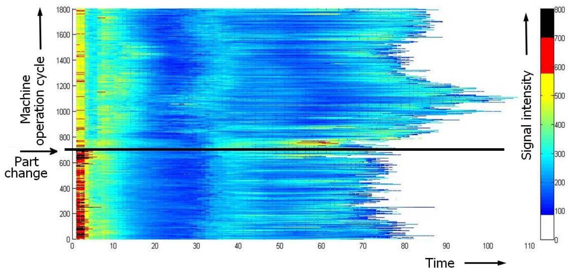

Fig. 4 Example of local changes in the cycles

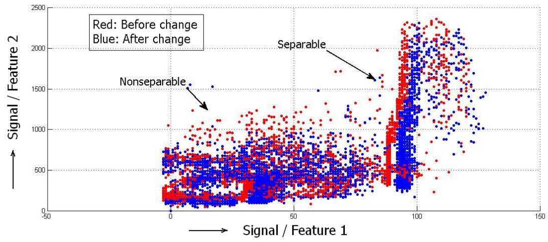

Fig. 5 Local classification indicators

step for further extraction of features. Furthermore, it is important for the purpose of study to divide this process into data segments. This is because of the nature of parameters of the process. The complete cycle of the machine comprises the process. The process comprises actions of moving cylinders, and these actions have different importance, different shape, and accordingly, different characteristics of related real signals. Additionally, these actions can be treated independently. By aid of control signals of the machine in addition to comparisons of signal values, the machine cycle is divided into 16 data segments (Fig. 6), each data segment has similar characteristics in all cycles. All these data segments of the machine process would give information about the behavior of the machine. However, some data segments are more informative and have more classification indicators than others. To recognize the segments which have more classification indicators the separable points in the feature space (Fig. 5) should be detected. The training data is used as time series sensor data to build a feature space where the points with high separability of the training data are detected and recognized in the sensor time series data where the best candidate segments are specified. Data segment 7 and data segments 12-14 are found to be the best candidates to be considered. The data segment 7 is the power stage where the machine is subjected to the highest stresses. It is the main stage which is responsible for deterioration because of the direct effect of the load on the wear rate of materials. On the other hand, data segments 12-14 comprise pulling back the cylinder without any load and this would allow for materials contact without disturbance of load changes in material and quantity. Both data segments, 7 and 12-14, where taken into consideration during the design of the system to select the most reliable data segment in diagnosis. The values of vibration velocity, vibration acceleration, the system hydraulic pressure, and displacement, along the data segments 7 and 12-14 respectively where taken as attributes in the classification process.

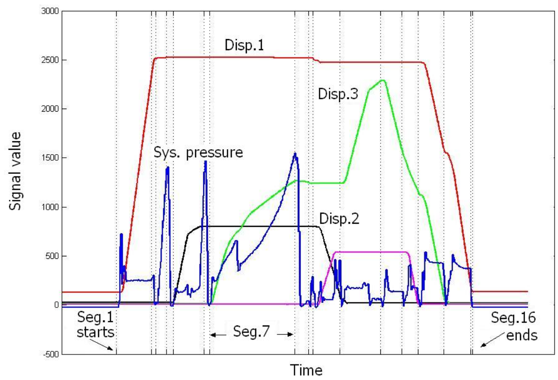

Fig. 6 An operation cycle of the machine

Tab. 1 Abbreviation

Parameter	Abbreviation
Vibration acceleration	A
Vibration velocity	V
System pressure	P
Displacement	D
No. of attributes	Attr.
No. of support vectors	S.V.
Accuracy	Acc.

4.2 Model selection

In SVM problems, model selection comprises selection of kernel, kernel parameters, and penalty parameter C. If the number of features is large, as in the current case, one may not need to map the data to a higher dimensional space, due to the nonlinear mapping which does not improve the performance, [9]. As observed in the current problem, using the linear kernel is good enough compared to other kernels used, and one only searches for the parameter C. A cross validation is done to estimate the best values of the penalty parameter C. The goal is to identify good value to C so that the unknown data (testing data) can be classified accurately, and not only the training data which comprises overfitting. Additionally, a simple scaling is done to the data to avoid numerical difficulties during calculation. Scaling is applied to help preventing the domination of greater numeric ranges attributes on the smaller numeric ranges ones, [9] .

Two classes are defined to train the classifier; the first one is the state $before$ wear part change, and the second is the state $after$ wear part change. A training data of

200 cycles, 100 cycles each class, were taken randomly from 4 places of the data. A linear kernel is considered because of the high number of attributes.

5 Results and discussion

Tab. 2 Results: Data segment 7 for 4.2004

Signal comb.	Attr.	S.V.	Acc. (Perc.)
A	100	20	56.94
V	100	11	60.71
P	100	14	56.15
D	100	9	54.03
A-V	200	19	60.46
A-P	200	14	56.49
A-D	200	10	56.06
P-V	200	9	58.79
V-D	200	9	57.76
P-D	200	9	54.47
A-P-V	300	16	58.59
A-V-D	300	9	56.87
A-P-D	300	10	55.98
P-V-D	300	9	57.71
A-P-V-D	400	9	56.84

Tab. 3 Results: Data segment 12-14 for 4.2004

Signal comb.	Attr.	S.V.	Acc. (Perc.)
A	200	22	94.54
V	200	21	96.96
P	200	70	91.56
D	200	63	93.66
A-V	400	25	97.36
A-P	400	24	95.53
A-D	400	21	96.06
P-V	400	21	97.14
V-D	400	22	97.34
P-D	400	52	95.41
A-P-V	600	26	97.44
A-V-D	600	25	97.67
A-P-D	600	26	96.23
P-V-D	600	22	97.44
A-P-V-D	800	25	97.76

Two groups of data are investigated: data segment 7 and data segment 12-14, with fifteen combinations of the four sets of measurement data. Each test set has a size 15874 cycles. Using the algorithms of Chang and Lin [10], the task was to solve the classification problem with a linear classifier, two classes, 200 cycles training data, 15874 cycles test data, and number of attributes 100, 200, 300, 400, 600, and 800 attributes.

Results of possible sensor combinations are summarized in table 2 and table 3 for data segment 7 and data segment 12-14 respectively and the abbreviations used are listed in table 1. In general, data segment 12-14 has better accuracy with maximum accuracy of 97.76 per-

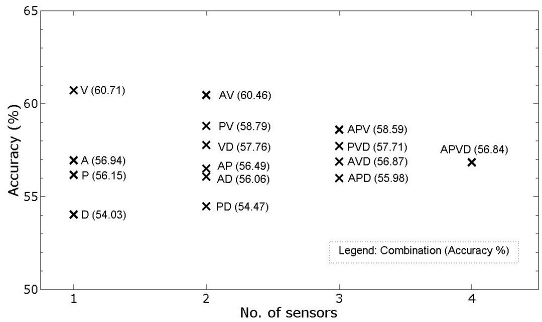

Fig. 7 Combination accuracies for seg. 7

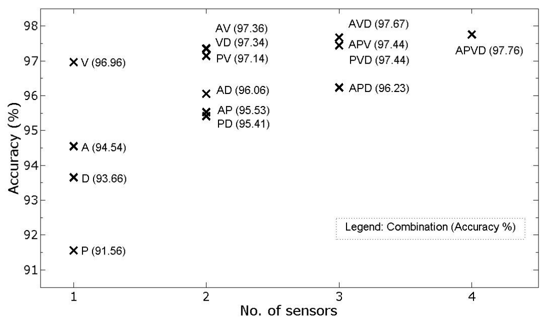

Fig. 8 Combination accuracies for seg. 12-14

cent than data segment 7 with maximum accuracy 60.71 percent. This is because data segment 12-14 comprises a sensor signal without disturbance from material type and size to be compressed.

In order to understand the effect of combining individual signals the possible combinations and their accuracies are plotted in the combination accuracies plots, Fig. 7 and Fig. 8 of segment 7 and segment 12-14 respectively. These plots help to recognize the effect of combining sensor information, whether positive or negative, on the accuracy of the classification system. They can also be used to follow a certain sensor and to realize the efficiency of adding different sensors to it. Studying the two figures, Fig. 7 and Fig. 8, the following observations can be realized: The best single signal accuracy in both groups is the results from the vibration velocity. This is may be because of the previously mentioned effect of wear frequency and the sensitivity of vibration velocity measurements to wear problems. The second best single signal accuracy is the vibration acceleration. It can also be seen from the two figures that the accuracy is negatively influenced by combining sensors in segment 7, while it is positively influenced by combining sensors in segment 12-14. This could be justified by the high level of noise in segment 7. Adding sensor information can lead to accumulating noise which would deteriorate the accuracy of classification. The best accuracy in both segments is the combination of all signals in data segment 12-14 with an accuracy of

97.76 percent. To avoid material-related disturbances the following aspects should be noted:

1. The higher dimensional feature space allows better description of the data.
2. The combination of signals can give better information than individual signals.
3. All sensor signals should be integrated so less information should be lost.

This is not the case with the results of segment 7 where the disturbance confused the results of the signal combinations.

To study the effect of combining a specified sensor with other sensors the vibration velocity is taken as an example to be followed in both segments. Accuracies of vibration velocity combinations are shown in Fig. 9 and Fig. 10 for segment 7 and segment 12-14 respectively. In addition to the observations and arguments mentioned with Fig. 7 and Fig. 8, it can be seen that combining a specified signal with a first signal of better accuracy than a second one gives better accuracy than combining the specified signal with the second signal.

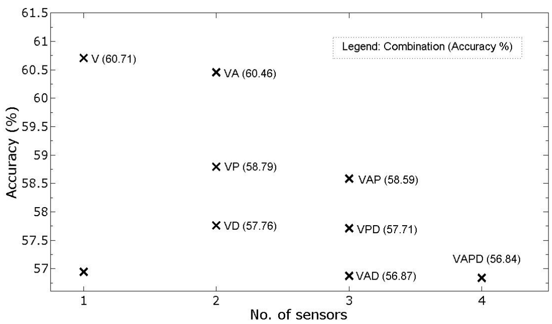

Fig. 9 Combination accuracies for vibration velocity in seg. 7

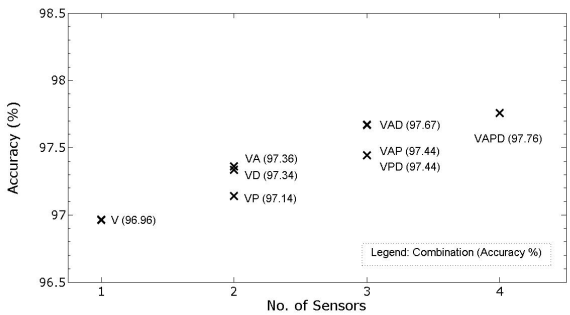

Fig. 10 Combination accuracies for vibration velocity in seg. 12-14

The decision value function of the highest accuracy combination is shown in Fig. 11. It is an indication of how much a cycle's information contributes to a class. Indeed it is a measure of how far a point (cycle) from the separating hyperplane. The running-in period of the new material can be clearly identified from the figure.

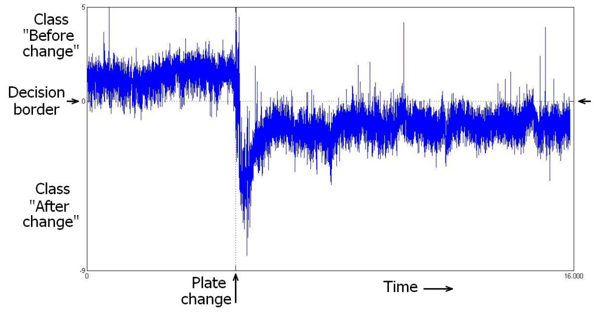

Fig. 11 Decision value of the signal combination for data segment 12-14

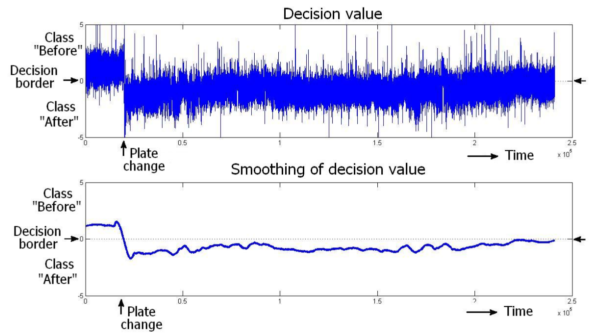

Fig. 12 Decision value of the extended test data and its smoothing

The model of the combination of signals in data segment 12-14 is applied to a longer test set data of 241034 cycles (average period of 1.5 year of machine operation). The resulting decision value function and its smoothing are shown in Fig. 12. Smoothing of the function shows a tendency for the average signal to increase over time. Briefly described, the two classes become closer over time. As a sequence, a change index (CI) is proposed based on the decision value function which shows a tendency to change over time coinciding with the deterioration of the part and the remaining life time. The CI, Fig. 13, is a measure function of the distance between the current position of the decision value of a measurement and the maximum allowed position of the decision value before changing of the wear part is necessary. The CI is not only an indication of how "worn" a part is, but it can give indication of the remaining life time of the part, which adds a value to the maintenance planning of the machine.

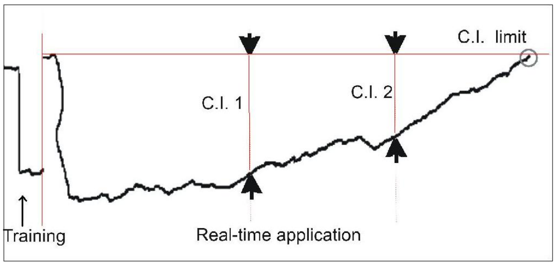

Fig. 13 The CI based on the decision value

6 Conclusion

A SVM classifier was applied to wear parts evaluation detection problems to find the most reliable characteristics for a diagnosis system. Combinations of sensor information are investigated to study the effect of fusing features on the accuracy of the classification. Application of SVM with preprocessing is efficient enough so that no further dimensional reduction is required. Vibration velocity can be detected as very sensitive to material changes, but nevertheless a combination of signals (including vibration velocity) gives more accurate result, with the friction occurring without disturbance of machine's operation process. The two states of running-in and wear-out of the considered parts can be distinguished clearly in the feature space, although they have some similar characteristics. Finally, a change index (CI) is proposed based on the decision function result of the classifier as an indication of how "worn" a part is.

7 References

[1] C. Cortes and V. Vapnik. Support vector networks. *AT & T Labs Research*, 1995.

[2] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. Technical report, University of Dortmund, 1997.

[3] O. Chapelle, P. Haffner, and V. Vapnik. Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, 10:1055–1064, 1999.

[4] M. Rychetsky, S. Ortmann, and M. Glesner. Support vector approaches for engine knock detection. *International joint conference on neural networks*, 2:969–974, 1999.

[5] C. Saunders, A. Gammerman, H. Brown, and G. Donald. Application of support vector machines to fault diagnosis and automated repair. *Eleventh international workshop on principles of diagnostics*, 2000.

[6] S. Abe. *Support vector machines for pattern classification*. Springer-Verlag, London, 2005.

[7] T. Huang, V. Kecman, and I. Kopriva. *Kernel-based algorithms for mining huge data sets*. Springer-Verlag, Berlin, 2006.

[8] Z. Yu G. Guohua, Z. Yongzhong and D. Guanghuang. Data fusion and multi-fault classification based on support vector machines. *The 9th Joint Conference on Information Sciences, Kaohsiung, Taiwan*, 2006.

[9] C. Hsu, C. Chang, and C. Lin. A practical guide to support vector classification. Technical report, National Taiwan University, 2008.

[10] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.