

Reliable bearing fault diagnosis using Bayesian inference-based multi-class support vector machines

M. M. Manjurul Islam, Jaeyoung Kim, Sheraz A. Khan, et al.

Citation: *The Journal of the Acoustical Society of America* **141**, EL89 (2017); doi: 10.1121/1.4976038

View online: <https://doi.org/10.1121/1.4976038>

View Table of Contents: <https://asa.scitation.org/toc/jas/141/2>

Published by the *Acoustical Society of America*

ARTICLES YOU MAY BE INTERESTED IN

[Improvement in thrust force estimation of solenoid valve considering minor hysteresis loop](#)

AIP Advances **7**, 056607 (2017); <https://doi.org/10.1063/1.4973388>

[Investigation of effects of long-term thermal aging on magnetization process in low-alloy pressure vessel steels using first-order-reversal-curves](#)

AIP Advances **7**, 056002 (2017); <https://doi.org/10.1063/1.4973605>

[Derivative measurement by frequency mixing](#)

Review of Scientific Instruments **47**, 1016 (1976); <https://doi.org/10.1063/1.1134829>

[Modeling of nanofabricated paddle bridges for resonant mass sensing](#)

Review of Scientific Instruments **77**, 073301 (2006); <https://doi.org/10.1063/1.2221560>

[A single-cell scraper based on an atomic force microscope for detaching a living cell from a substrate](#)

Journal of Applied Physics **118**, 134701 (2015); <https://doi.org/10.1063/1.4931936>

[Effects of the computational domain size on direct numerical simulations of Taylor-Couette turbulence with stationary outer cylinder](#)

Physics of Fluids **27**, 025110 (2015); <https://doi.org/10.1063/1.4913231>

JASA
THE JOURNAL OF THE
ACOUSTICAL SOCIETY OF AMERICA

Special Issue:
Additive Manufacturing and Acoustics

Read Now!

Reliable bearing fault diagnosis using Bayesian inference-based multi-class support vector machines

M. M. Manjurul Islam, Jaeyoung Kim, Sheraz A. Khan, and
Jong-Myon Kim^{a)}

School of Electrical, Electronics, and Computer Engineering, University of Ulsan, Ulsan,
South Korea
m.m.manjurul@gmail.com, kjy7079@gmail.com, sherazalik@gmail.com,
jongmyon.kim@gmail.com

Abstract: This letter presents a multi-fault diagnosis scheme for bearings using hybrid features extracted from their acoustic emissions and a Bayesian inference-based one-against-all support vector machine (Bayesian OAASVM) for multi-class classification. The Bayesian OAASVM, which is a standard multi-class extension of the binary support vector machine, results in ambiguously labeled regions in the input space that degrade its classification performance. The proposed Bayesian OAASVM formulates the feature space as an appropriate Gaussian process prior, interprets the decision value of the Bayesian OAASVM as a maximum *a posteriori* evidence function, and uses Bayesian inference to label unknown samples.

© 2017 Acoustical Society of America

[DHC]

Date Received: October 17, 2016 **Date Accepted:** January 5, 2017

1. Introduction

Rolling element bearings are the most frequently failing components in rotating machinery, accounting for approximately 51% of all failures (Kang *et al.*, 2015a). Bearing failure is generally caused by localized defects, which appear as cracks or spalls on rollers or raceways (inner or outer) of a bearing (Kang *et al.*, 2015b). Data-driven techniques are extensively used for fault diagnosis in bearings. These methods work by means of three steps: data or signal acquisition, extraction of features from the acquired data, and classification of the data based upon the extracted features. Generally, vibration and current signals are utilized to detect faults in bearings, but the analysis of these signals is not effective in detecting emergent faults at low operating speeds (Widodo *et al.*, 2009). In this letter, we describe a data-driven method based upon acoustic emission (AE) signals to detect multiple types of single and compound bearing defects in machines operating at low speeds. The diagnostic performance of this method is improved using a Bayesian one-against-all support vector machine (Bayesian OAASVM).

Statistical features calculated for the time- and frequency-domain AE signal, along with features extracted through complex envelope analysis of the AE signal, are used to create a hybrid feature vector. A hybrid feature vector is useful in accurately identifying each fault condition; nevertheless, the ultimate diagnostic performance largely depends upon the effectiveness of the classifier. Generally, classifiers like naive Bayes, artificial neural networks, and support vector machines (SVMs) (Widodo *et al.*, 2009; Islam *et al.*, 2015) are used to generate models of the training data, which are then used to classify unknown test data. SVM is the most extensively used method because of its better generalization performance and its ability to work well with high-dimensional input data (Widodo *et al.*, 2009).

The standard one-against-all support vector machine (standard OAASVM) is the most widely used multi-class extension of the original binary SVM. It constructs l binary SVMs for an l -class classification problem, where the k th SVM is used to distinguish class k from the remaining $l - 1$ classes. An unknown feature vector is classified only if it is accepted by one of the l binary SVMs and rejected by the remaining $l - 1$. However, this is not always the case; a feature vector might be rejected by all the SVMs or accepted by more than one SVM, resulting in ambiguously labeled regions of the input space and hence degradation in diagnostic performance. The

^{a)} Author to whom correspondence should be addressed.

Table 1. Summary of AE data acquisition conditions, including the use of two different operating conditions and two crack sizes.

| Datasets ^a | Average rotational speed (RPM) | Sizes of cracks in the bearing's outer and/or inner roller raceways | | |
|-----------------------|--------------------------------|---|---------|---------|
| | | Length | Width | Depth |
| Dataset 1 Dataset 2 | 300 500 | 3 mm | 0.35 mm | 0.30 mm |
| Dataset 3 Dataset 4 | 300 500 | 12 mm | 0.49 mm | 0.50 mm |

^aNinety AE signals for each fault type; sampling frequency $f_s = 250$ kHz; each signal is 10 s long.

“fuzzy one-against-all support vector machines” (fuzzy OAASVMs) classifier improves the accuracy of the standard OAASVM by calculating class membership values for samples in the ambiguously labeled regions (Abe, 2015). Likewise, Islam et al. (2015) have improved the standard OAASVM by assigning a static reliability measure to individual SVMs and by proposing a new decision aggregation rule (Islam et al., 2015). None of these methods, however, use the SVM in a probabilistic framework to estimate the likelihood of an unknown observation being a member of class; rather, they use it only to obtain class label information.

In this letter, we propose the Bayesian multi-class one-against-all support vector machine (Bayesian OAASVM), which considers the standard OAASVM as a maximum *a posteriori* evidence function based on the appropriate formulation of the feature space as a Gaussian process prior (GPP), and then estimates the class probabilities of the unknown samples using the principles of Bayesian inference (Murphy, 2012). The proposed Bayesian OAASVM is used to improve the diagnostic performance of fault diagnosis schemes in low-speed rotary machines using AEs.

2. Acquisition of AE data

We use a previously reported experimental setup (Islam et al., 2015; Kang et al., 2015b) to capture AEs from normal bearings and bearings with seeded defects, both single and compound. The AEs are captured using a wideband AE sensor and sampled at 250 kHz using a PCI-2 system. AE signals are recorded for bearings operating at two different rotational speeds with seeded defects of two different dimensions. A total of four datasets are analyzed, as summarized in Table 1. Each dataset has signals for eight bearing conditions: normal condition (BNC), outer raceway crack (BCO), inner raceway crack (BCI), roller crack (BCR), inner and outer raceway cracks (BCIOs), outer and roller cracks (BCORs), inner and roller cracks (BCIRs), and inner, outer, and roller cracks (BCIORs).

3. Method for fault diagnosis in bearings

The proposed method for fault diagnosis consists of hybrid feature extraction and classification using the proposed Bayesian OAASVM; Fig. 1 illustrates the method in detail.

3.1 Hybrid feature extraction

Data-driven techniques detect bearing defects by using different features of the fault signal. The accurate detection of multiple bearing defects requires the extraction of distinguishing features from the AE signal that can be used to uniquely identify each

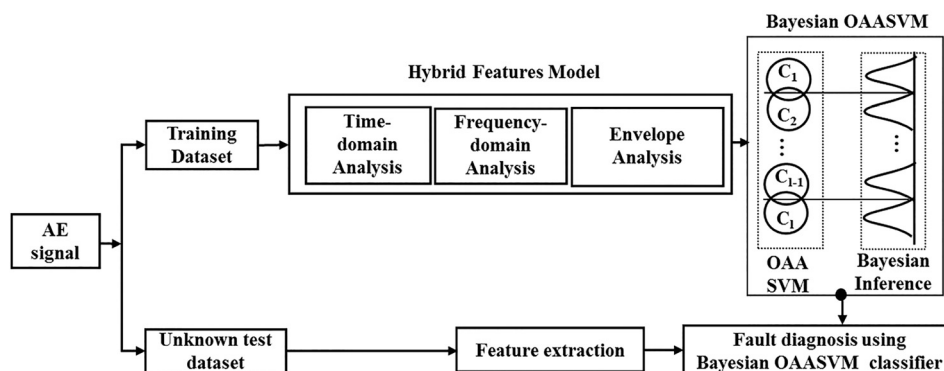


Fig. 1. Detailed framework of the reliable bearing fault diagnosis scheme. “C” represents the fault classes.

defect. We use a hybrid feature model that requires the calculation of different statistical measures of the time and frequency domain AE signal, as well as the signal's envelope power spectrum. Twelve statistical features of the time and frequency domain AE signal are calculated as given in Table 2 (Kang et al., 2015a). Moreover, 12 features are extracted from the envelope power spectrum of the AE signal, including root-mean-square (RMS) values for the three characteristic defect frequencies, and the first three harmonics of each of these frequencies. The characteristic defect frequencies are the ball pass frequency over inner race, ball pass frequency over outer race, and two times the ball spin frequency; the envelope power spectrum of the AE signal shows peaks at these frequencies and their harmonics for defective bearings (Kang et al., 2015b). For each AE signal, feature vectors are constructed using these 24 features, which are then used to train the Bayesian OAASVM that is subsequently used to classify unknown AE signals.

3.2 Proposed Bayesian OAASVM

Consider an l -class classification problem with the dataset $Q = \{(x_i, y_i) | x_i \in R^d\}_{i=1}^n$, where $x_i \in R^d$ is a d -dimensional feature vector, $y_i \in \{1, 2, \dots, l\}$ is the set of class labels, and n is the number of feature vectors in the training dataset. In the standard OAASVM, the following optimization problem is solved to distinguish a particular class $k = 1$ from the remaining $l - 1$ classes (Chih-Wei and Chih-Jen, 2002)

$$\begin{aligned} & \text{minimize} \left\{ \frac{1}{2} \|\omega_i\|^2 + C \sum_{j=1}^n (y_j [\omega \cdot \varphi(x_j) + b]) \right\} \\ & \text{subject to} \quad y_j (\omega_i \cdot \varphi(x_j) + b_i) \geq 1 - \zeta_j^i, \text{ if } y_j = i \\ & \text{and} \quad \zeta_j^i \geq 0, \forall i = 1, 2, \dots, n. \end{aligned} \quad (1)$$

Here, b is the bias, ω is the weight vector, $\varphi(x_j)$ is the kernel function that maps input feature vectors x_j to a high-dimensional space, where they are linearly separable by a hyperplane with a maximum margin of $b/\|\omega\|$, and C is the linearity constraint. During classification, the standard OAASVM labels a feature vector x as i^* if the decision function f_i generates the highest value for i^* , as given in Eq. (2),

$$i^* = \operatorname{argmax}_{i=1,2,\dots,l} f_i(x) = \operatorname{argmax}_{i=1,2,\dots,l} (\omega_i^T \varphi(x) + b_i). \quad (2)$$

Moreover, the value of i^* th decision function should be positive, and the values of the remaining decision functions should be negative as given in Eq. (3),

$$f_{i=i^*}(x) > 0, \quad f_{i \neq i^*}(x) < 0. \quad (3)$$

The feature vectors that do not satisfy the criterion in Eq. (3) are not classified by the standard OAASVM and are defined as *ambiguous feature vectors*, as follows:

$$\forall \bar{x} \notin \{x | f_{i=i^*}(x) > 0, \quad f_{i \neq i^*}(x) < 0\}. \quad (4)$$

These ambiguous feature vectors are classified using Bayes' rule. The conditional probability of the feature vector x being labeled as i , is given in Eq. (4),

$$p(y_i | x_1, \dots, x_d) = \frac{p(x_1, \dots, x_d | y_i) p(y_i)}{p(x)}, \quad \forall i = 1, 2, \dots, n, \quad (5)$$

where $p(x)$ is the probability of the feature vector x , $p(x_1, \dots, x_d | y_i)$ is the conditional probability of feature vector x given class label y_i , and $p(y_i)$ is the probability of class y_i . The probability $p(y_i)$ is equal to $1/n$, as class y_i is selected among n classes. Since each feature vector is independent, the conditional probability, $p(x_1, \dots, x_d | y_i)$, can be written as follows:

$$p(x_1, \dots, x_d | y_i) p(y_i) = p(x_1 | y_i) p(x_2 | y_i), \dots, p(x_d | y_i) p(y_i) = p(y_i) \prod_{k=1}^d p(x_k | y_i). \quad (6)$$

Finally, the conditional probability $p(y_i | x_1, \dots, x_d)$ can be determined as follows:

$$p(y_i | x_1, \dots, x_d) = \frac{p(y_i) \prod_{k=1}^d p(x_k | y_i)}{p(x_1, \dots, x_d)}, \quad (7)$$

Table 2. Six time-domain and three frequency-domain statistical feature components of the AE signal (s is the time domain of the signal, f is the frequency domain of the s signal) with their equations.

| Features | Equations | Features | Equations | Features | Equations | Features | Equations |
|--|--|--------------------------|--|-----------------------|--|-----------------|---|
| Time-domain: RMS | $\left(\frac{1}{N} \sum_{i=1}^N s_i^2\right)^{1/2}$ | Square root of magnitude | $\left(\frac{1}{N_{\text{sample}}} \sum_{n=1}^{N_{\text{samples}}} \sqrt{ s(n) }\right)^2$ | Skewness | $\frac{1}{N} \sum_{i=1}^N \left(\frac{s_i - \bar{s}}{\sigma}\right)^3$ | Crest factor | $\frac{\max(s_i)}{\text{RMS}}$ |
| Kurtosis | $\frac{1}{N} \sum_{i=1}^N \left(\frac{s_i - \bar{s}}{\sigma}\right)^4$ | Shape factor | $\frac{\text{RMS}}{\left(\frac{1}{N} \sum_{i=1}^N s_i \right)}$ | Impulse factor | $\frac{\max(s_i)}{\frac{1}{N} \sum_{i=1}^N s_i }$ | Kurtosis factor | $\frac{\text{Kurtosis}}{\left(\frac{1}{N} \sum_{i=1}^N s_i^2\right)^2}$ |
| Frequency-domain: RMS frequency | $\left(\frac{1}{N} \sum_{i=1}^N f_i^2\right)^{1/2}$ | Root variance frequency | $\left(\frac{1}{N} \sum_{i=1}^N (f_i - \text{FC})^2\right)^{1/2}$ | Frequency center (FC) | $\frac{1}{N} \sum_{i=1}^N f_i$ | | |

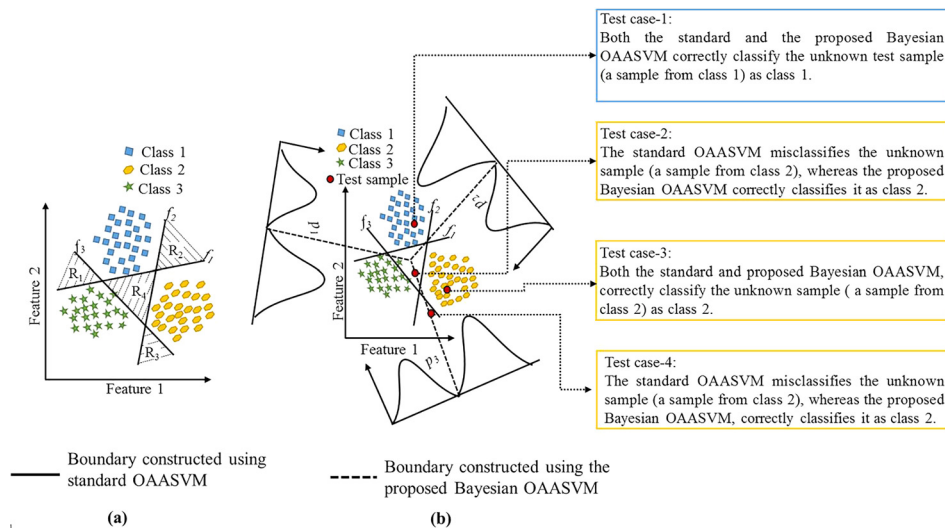


Fig. 2. (Color online) (a) Illustration of the problem of standard OAASVM for a 3-class classification problem, in which the use of the standard decision functions leads to overlapped classification regions ($R_1, R_2, R_3,$ and R_4). (b) Resolution of this problem using the proposed probabilistic decision function.

where $p(x_1, \dots, x_d)$ is equal to 1, as x is an observation. For estimating $p(x_k|y_i)$, we assume that the probability distribution of feature x_k is Gaussian, and hence $p(x_k|y_i)$ is calculated using the training data, as follows:

$$p(x_k|y_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(x_k-\mu_i)^2/2\sigma_i^2}, \tag{8}$$

where μ_i and σ_i^2 are the mean and standard deviations of the feature value in the i th class, respectively. For classification, an ambiguous feature vector \bar{x} is labeled as i^* , if the conditional probability $p(y_i|\bar{x}_1, \dots, \bar{x}_d)$ is highest for i^* , as given in Eq. (9),

$$i^* = \underset{i}{\operatorname{argmax}}(p(y_i|\bar{x}_1, \dots, \bar{x}_d)). \tag{9}$$

In Fig. 2, we consider a 3-class classification problem to illustrate the effectiveness of the proposed Bayesian OAASVM, which uses the probabilistic decision function in Eq. (9) as opposed to the decision function in Eq. (2) that is employed in the standard OAASVM. The standard OAASVM fails to correctly classify the data points in the overlapped regions, whereas the Bayesian OAASVM correctly classifies these data points.

Table 3. Average sensitivities of the proposed model and other models for each fault type and each dataset.

| Datasets | Average sensitivity for each fault type | | | | | | | | | |
|-----------|---|--------|--------|--------|--------|--------|--------|-------|--------|----------|
| | OAASVM | BCO | BCI | BCR | BCIO | BCOR | BCIR | BCIOR | BNC | Avg. (%) |
| Dataset 1 | Standard | 87.57 | 85.84 | 77.10 | 78.37 | 81.44 | 86.24 | 83.00 | 83.11 | 82.83 |
| | Fuzzy | 85.00 | 93.00 | 96.33 | 93.00 | 95.00 | 90.55 | 91.00 | 90.43 | 91.79 |
| | Bayesian | 93.78 | 100.00 | 100.00 | 97.33 | 96.89 | 100.00 | 97.78 | 100.00 | 98.22 |
| Dataset 2 | Standard | 80.66 | 82.13 | 82.66 | 82.53 | 82.53 | 82.39 | 82.26 | 82.21 | 82.17 |
| | Fuzzy | 89.28 | 93.08 | 94.62 | 89.48 | 96.22 | 91.35 | 90.22 | 93.97 | 92.28 |
| | Bayesian | 100.00 | 100.00 | 99.56 | 100.00 | 98.67 | 100.00 | 98.22 | 100.00 | 99.55 |
| Dataset 3 | Standard | 91.69 | 92.00 | 93.69 | 92.49 | 82.09 | 90.36 | 91.96 | 96.00 | 91.29 |
| | Fuzzy | 91.9 | 95.1 | 83.03 | 93.57 | 98.3 | 92.77 | 93.03 | 95.43 | 92.89 |
| | Bayesian | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| Dataset 4 | Standard | 92.31 | 93.92 | 93.12 | 92.22 | 93.31 | 92.17 | 98.00 | 93.30 | 93.54 |
| | Fuzzy | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 96.00 | 92.59 | 99.50 | 98.51 |
| | Bayesian | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

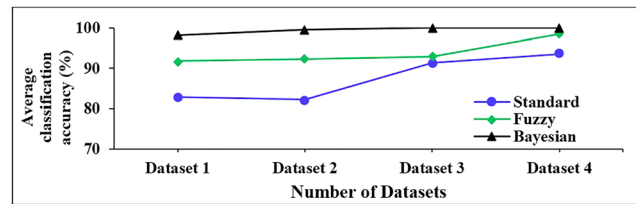


Fig. 3. (Color online) Average classification accuracies of the proposed model and other state-of-the-art models.

4. Experimental results and discussion

The proposed Bayesian OAASVM classifier is validated by using it in a multi-class bearing diagnostics classification scenario, which considers four datasets with eight fault classes in each dataset (see Table 1). Each dataset contains 90 AE signals for each fault condition, and these are randomly divided into training and test sets with 40 and 50 AE signals, respectively. For each AE signal, a hybrid feature vector is extracted as discussed in Sec. 3.1. These hybrid feature vectors are then used as inputs to the standard, fuzzy, and proposed Bayesian OAASVMs. The performance of the Bayesian OAASVM is compared with the standard OAASVM and the fuzzy OAASVM in terms of sensitivity and average classification accuracy. Table 3 presents the experimental results. The proposed Bayesian OAASVM delivers better classification performance than the other two approaches, yielding classification accuracy of 98.22%, 99.55%, 100%, and 100% for datasets 1 to 4, respectively (Fig. 3). The proposed Bayesian OAASVM improves the average classification accuracy of the diagnostic system by 17.83% and 4.68% as compared to the standard OAASVM and the fuzzy OAASVM, respectively. The Bayesian OAASVM improves the diagnostic performance of the bearing fault diagnosis scheme by more accurately labeling the ambiguous feature vectors or samples in the overlapped regions of the input space.

5. Conclusions

In this letter, we proposed a Bayesian OAASVM classifier to improve the diagnostic performance of a multi-class bearing fault diagnosis scheme. The fault diagnosis scheme used hybrid feature vectors extracted from the AEs of normal and defective bearings. The proposed Bayesian OAASVM improved the classification accuracy of the multi-class fault diagnosis scheme by accurately labeling feature vectors in the ambiguously labeled regions of the input space. Ambiguously labeled regions in the input space are a common outcome of extending the original binary SVM to multi-class classification problems. The Bayesian OAASVM uses a GPP to properly utilize the feature space, and Bayesian inference to correctly label feature vectors in the ambiguously labeled regions. The proposed Bayesian OAASVM yielded superior diagnostic performance compared to the standard OAASVM and fuzzy OAASVM. The overall improvement in average classification accuracy ranged from 4.68% to 17.83% across different datasets.

Acknowledgments

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (Grant Nos. 20162220100050 and 20161120100350), in part by the Leading Human Resource Training Program of Regional Neo industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and future Planning (Grant No. NRF-2016H1D5A1910564), in part by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2016 (Grant Nos. C0395147 and S2381631), in part by Basic Science Research Program through the NRF funded by the Ministry of Education (Grant No. 2016R1D1A3B03931927).

References and links

- Abe, S. (2015). "Fuzzy support vector machines for multilabel classification," *Pattern Recogn.* **48**, 2110–2117.
- Chih-Wei, H., and Chih-Jen, L. (2002). "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Networks* **13**, 415–425.
- Islam, M. M. M., Khan, S. A., and Kim, J.-M. (2015). "Multi-fault diagnosis of roller bearings using support vector machines with an improved decision strategy," in *Proceedings of the Advanced Intelligent Computing Theories and Applications: 11th International Conference, ICIC 2015, Part III*, Fuzhou, China

(August 20–23, 2015), edited by D.-S. Huang and K. Han (Springer International Publishing, Cham), pp. 538–550.

Kang, M., Kim, J., Kim, J. M., Tan, A. C. C., Kim, E. Y., and Choi, B. K. (2015a). “Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis,” *IEEE Trans. Power Electron.* **30**, 2786–2797.

Kang, M., Kim, J., Wills, L. M., and Kim, J. M. (2015b). “Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis,” *IEEE Trans. Indust. Electron.* **62**, 7749–7761.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective* (MIT Press, Cambridge, MA), pp. 290–295.

Widodo, A., Kim, E. Y., Son, J.-D., Yang, B.-S., Tan, A. C. C., Gu, D.-S., Choi, B.-K., and Mathew, J. (2009). “Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine,” *Expert Syst. Appl.* **36**, 7252–7261.