

# Wind Turbine Blade Damage Detection Using Supervised Machine Learning Algorithms

**Taylor Regan**

Department of Mechanical Engineering,  
University of Massachusetts Lowell,  
Lowell, MA 01854

**Christopher Beale**

Department of Mechanical Engineering,  
University of Massachusetts Lowell,  
Lowell, MA 01854

**Murat Inalpolat<sup>1</sup>**

Department of Mechanical Engineering,  
University of Massachusetts Lowell,  
1 University Avenue,  
Lowell, MA 01854  
e-mail: Murat\_Inalpolat@uml.edu

*Wind turbine blades undergo high operational loads, experience variable environmental conditions, and are susceptible to failure due to defects, fatigue, and weather-induced damage. These large-scale composite structures are fundamentally enclosed acoustic cavities and currently have limited, if any, structural health monitoring (SHM) in place. A novel acoustics-based structural sensing and health monitoring technique is developed, requiring efficient algorithms for operational damage detection of cavity structures. This paper describes the selection of a set of statistical features for acoustics-based damage detection of enclosed cavities, such as wind turbine blades, as well as a systematic approach used in the identification of competent machine learning algorithms. Logistic regression (LR) and support vector machine (SVM) methods are identified and used with optimal feature selection for decision-making via binary classification algorithms. A laboratory-scale wind turbine with hollow composite blades was built for damage detection studies. This test rig allows for testing of stationary or rotating blades, of which time and frequency domain information can be collected to establish baseline characteristics. The test rig can then be used to observe any deviations from the baseline characteristics. An external microphone attached to the tower will be utilized to monitor blade health while blades are internally ensonified by wireless speakers. An initial test campaign with healthy and damaged blade specimens is carried out to arrive at several conclusions on the detectability and feature extraction capabilities required for damage detection.*  
[DOI: 10.1115/1.4036951]

*Keywords: wind turbine blades, health monitoring, machine learning, support vector machine, logistic regression*

## 1 Introduction

As wind energy increases its industry market share, wind farm operators are investigating ways to decrease operation and maintenance costs [1]. Field experience has determined that blades are the highest contributors to wind turbine maintenance costs [2,3]. As composite turbine blades continue to increase in size (improving energy output), it becomes harder to retain structural integrity in operation, which in turn amplifies the cost associated with operation and maintenance. Turbine blades are subjected to aerodynamic and gravitational loads under varying environmental conditions, which can result in cracking, holes, delamination, and deformation. While the blades are traditionally inspected using a time-based maintenance strategy, these inspections are relatively infrequent, expensive, pose a physical risk to the inspector, and are reliant on what can be visually detected by the inspector [4]. Developing a smart methodology for condition-based maintenance would stand to significantly reduce costs.

Several inspection methods are currently available for detecting various blade faults such as holes, cracks, and delamination. Vibration, acoustic emission, and wave propagation-based methods are frequently used in order to evaluate the structural integrity of turbine blades [5–7]. Photogrammetry-based optical techniques have recently been used to detect defects in wind turbine blades [8,9]. Ultrasound-based inspection is one of the most widely used blade inspection techniques in industry and has been successfully demonstrated in the literature [10–12]. Acoustic beamforming is a

versatile but costly technique that has been demonstrated to succeed under certain conditions [13,14]. Acoustic signatures taken from the structure, before and after damage, may allow for detection and differentiation of fault existence and can also be used as a blade inspection method [14–18]. Aizawa et al. investigated damage detection of wind turbine blades by installing a speaker inside of a stationary wind turbine blade and qualitatively characterizing the sound radiation using a microphone array [14]. They ultimately observed that faults would change acoustic energy radiated from the test object. Arora et al. used vibroacoustic modal analysis to determine that, when exciting a structure via internal loudspeaker, vibroacoustically coupled mode shapes will change due to damage on the structure [5].

In addition to the numerous proposed inspection methods, there are a number of data processing schemes proposed to arrive at conclusions about the health of the turbine blades. Nair and Kiremidjian use Gaussian mixture models as clustering algorithms in pattern classification. Initial parameters are estimated through the use of the  $K$ -means algorithm [19]. Sohn et al. utilize linear/quadratic discrimination methods in addition to statistical control charts to arrive at a damage diagnosis [20]. Principal component analysis is a well-established technique that has been used to map a large set of input features to a much smaller combination of features (which are linear combinations of the original ones) [21]. Krause et al. use a model of a cracking sound in comparison with acoustic detection of a structure [22]. They use five features extracted from the comparison: power slope, tonality, spectrum slope, spectrum similarity, and impulse decay, to arrive at conclusions on the health state of the structure. Edwards et al. [23] propose a robust SHM system that uses an initial time-series algorithm (trained to predict system response using “baseline” data) to compare predicted response to measured response and

<sup>1</sup>Corresponding author.

Contributed by the Technical Committee on Vibration and Sound of ASME for publication in the JOURNAL OF VIBRATION AND ACOUSTICS. Manuscript received June 10, 2016; final manuscript received May 26, 2017; published online August 2, 2017. Assoc. Editor: Patrick S. Keogh.

generate a specified “damage indicator.” Furthermore, they use info-gap decision-making theory to measure the uncertainty and obtain the bounds of variation.

Autoregressive models in conjunction with time-series analysis have been widely used in SHM both for feature extraction as well as damage detection [24]. Figueiredo et al. use an autoregressive model to perform feature extraction from accelerometer time data measured on a structure under varying health conditions [25]. Auto-associative neural networks, factor analysis, Mahalanobis squared distance, and singular value decomposition algorithms were used to normalize data and generate a scalar damage indicator. Algorithm performance was evaluated using receiver operating characteristic curves [25]. Nick et al. used unsupervised learning on acoustic emission signals to identify the existence and location of damage, and then switched to supervised learning (support vector machines, naïve Bayes classifiers, feedforward neural networks, random forest ensemble learning, and AdaBoost) to identify the type and severity of damage [26]. They found SVMs to be the best performing, both in precision as well as classification time. Neural networks have widely been used for damage identification using vibration data, and support vector machines are also becoming increasingly utilized [27].

Most of the previously mentioned inspection methods are not applicable during wind turbine operation and are instead utilized for inspection when the blades are stationary. There is a clear need for the development of a new technique which will detect damage and ensure the integrity of wind turbine blades in operation. Niezrecki and Inalpolat have proposed a novel acoustic sensing-based structural health monitoring technology for use on operating turbine blades. The preliminary computational and experimental results of this method have been presented in earlier publications [28–30]. There is also a need for additional literature with regard to implementing machine learning algorithms to monitor structures via selected feature sets to observe and detect damage [20].

This paper extends the method of active acoustic detection by considering experimental data from a laboratory wind turbine test rig. The executed test plan addresses a variety of wind turbine blade damage types, locations, and severities. The blades are stationary throughout each test case, with the exception of one test case to include the influence of rotational blades on the data. Multiple acoustic excitation types, including single-tone harmonic, multitone harmonic, and white noise, are considered for each test case. Experimental results are evaluated using supervised machine learning techniques in order to interpret the health state of wind turbine blades in operation. The data presented include the laboratory measurements required for the proof-of-concept studies, with a focus on data interpretation via machine learning to observe and detect structural damage. This paper is expected to contribute to structural health monitoring methodologies and its application to practical systems such as wind turbines. Specifically, the contributions are expected to be in the fields of signal acquisition and processing, data interpretation, decision-making, and algorithm development for operational damage detection.

## 2 Methodology

**2.1 Acoustics-Based Blade Monitoring Methodology.** The intent of this study is to further develop an acoustics-based damage detection methodology by employing machine learning algorithms to determine whether maintenance is required on a turbine blade in operation. It is anticipated that the blade damage will manifest itself in changes to the statistical features extracted from the acoustic data.

The acoustic sensing method proposed by Inalpolat et al. [28–30] involved two approaches for discrimination of healthy and damaged specimens: passive and active detection. Passive detection seeks to determine damage by measuring the acoustic response of the blade cavity to ambient external wind flow. For active detection, the cavity is excited via an internally mounted speaker (with controlled output frequency and volume level). The sound radiation from the cavity is measured via a microphone mounted externally on the tower, which will allow cracks and other damage types to be observed. Results presented in this paper are derived from experimentation via the second method, “active detection.” The proposed active damage detection method is depicted in Fig. 1.

Active detection was used in this study as it allowed the excitation tones to be variables in the algorithm’s ability to discern output classes. Additionally, it allowed each blade to be ensounded with a separate unique tone, for later use in attempting to identify which blade is in possession of a fault, and to determine if the optimal feature vector changed between single-tone and multitone tests. Once a damaged blade is discovered, further testing will be performed on that blade, drawing more specific conclusions as to the health of the blade. The initial goal of this study is to use supervised machine learning techniques to identify whether damage is present in the blades. Using cavity acoustics and supervised machine learning techniques will provide a means of continuous structural health monitoring, even while blades are in operation, providing a more cost-effective and proactive monitoring scheme.

A concern regarding SHM implementation in industry is the volume of data accrued from continual collection of sound pressure information from the blades. Processing this data manually would require a trained individual to visually monitor all of the incoming data from each of the blades in a wind farm, quickly becoming prohibitively costly. To combat this, an acquisition strategy is proposed, which samples the blade’s acoustic signature at predetermined intervals (for example, every 30 mins). Each time signature is then condensed to a set of features which seeks to preserve as much information as possible from the time data while reducing the data to a much smaller dimension. Once the feature vector has been generated, machine learning algorithms will be used to process the information and diagnose/monitor the health of the blade.

**2.2 Feature Selection.** Proper feature selection is critical to the success of the health monitoring algorithm and involves identifying a feature set that contains as much information as possible about the system while minimizing the likelihood of false alarms

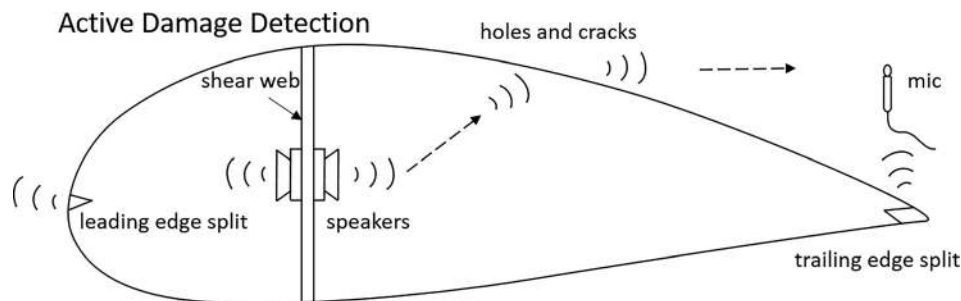


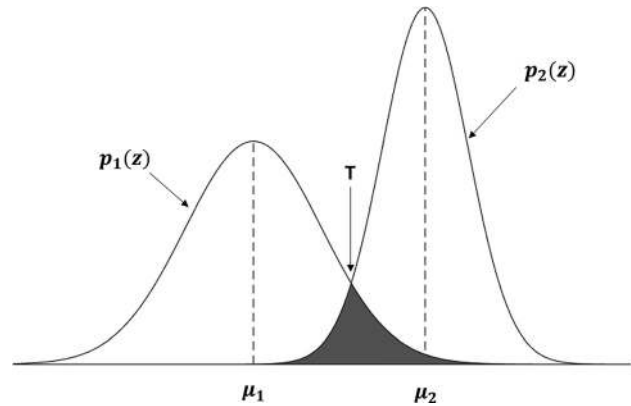
Fig. 1 Schematic of the active damage detection

[27,31]. In preparing the data for conversion to a set of representative features, it is good practice to first attenuate noise and other undesired signal artifacts. Selected features are often chosen based on previous experience and observation of fault-sensitive characteristics. Understanding how the system is expected to change due to damage is often helpful in selecting features. For example, faults can result in changes to the power spectrum, signal magnitude, and the energy of the system [31]. Good features are insensitive to operational and environmental variability, and can distinctly separate classes [31]. One method for determining which features are sensitive to the development of faults is to create damages on the structure that are similar to those anticipated to occur during the operational life of the component [32]. Poor selection of features can result in the variations due to environmental and operational conditions overwhelming any changes present due to actual damage. This problem occurs both in statistical approaches and model-based approaches [27].

A total of 12 features are considered in this study consisting primarily of signal statistics. A complete list of the features is as follows: mean, median, root-mean-square (RMS), root-sum-of-squares (RSSQ), mean frequency, median frequency, kurtosis, variance, crest factor, standard deviation, skewness, and peak amplitude of the fast Fourier transform (FFT). Signal statistics have been used as damage-sensitive features with success and can be computed with minimal computational effort and time [32]. All statistics that are considered are readily available functions in MATLAB, where all processing and algorithms were developed. Adding features requires more computational time, memory, and effort to process all the data. Limiting the number of features allowed the study to focus on proper evaluation of the machine learning algorithms and feature contribution metrics.

**2.3 Feature Contribution Metrics.** An alternative to selecting a new feature set for each unique analysis or running an algorithm with an unnecessarily large global feature vector is to develop a metric to establish how well each feature represents the data and select a subset of the global feature vector to represent the system. An automated procedure can select this subset prior to training the algorithm, reducing the time it takes to optimize the algorithm. In practice, feature vectors can become quite large. Data are collected from a myriad of sensors, in many locations, which can result in hundreds and thousands of processed features. In some applications, not all of these features are needed at the same time, and in machine learning, there is a balance between the number of features needed and the number of training examples needed. In terms of the efficiency of the damage detection algorithm, the best performance is often attained when limiting the feature vector to the features that are most sensitive to damage. Using a large input feature vector with features that are not very sensitive to damage takes a longer time to converge on an optimized algorithm solution than fewer features that are more fault-sensitive. Though the feature set presented in this paper is comparatively small (roughly a dozen features), two proposed feature contribution metrics are tested as a proof-of-concept study for implementation as an optimized feature selection methodology in an SHM system. In this paper, the distinguishability measure and Fisher's ratio are proposed as feature contribution metrics. These methods are selected due to ease of implementation and quick processing time. It is desirable that the optimized feature selection is a quick process, so that in real time, the training algorithm can determine which subset of an overall large feature vector is best applied under certain conditions to determine and monitor the health of the structure.

**2.3.1 Distinguishability Measure.** A feature's ability to differentiate between multiple classes can be quantified by its distinguishability. Figure 2 displays two probability distributions ( $p_1(z)$  and  $p_2(z)$ ), where  $z$  represents the values of a selected feature), plotted along the same axis, representative of two classes of some feature,  $A$ . It should be noted that both distributions are



**Fig. 2 Illustration of the feature distinguishability metric**

assumed to be Gaussian, which is valid for most of the large training datasets [31].

Distinguishability is a measure of how much overlap exists between two classes of a feature. In Fig. 2, this is symbolically represented as the shaded area. Intuitively, the smaller this area is, the better that particular feature is at distinguishing the classes of data. The distinguishability measure for a feature,  $A$ , is given in the following equation:

$$d_A^{w_1 w_2} = \int_{-\infty}^T \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(z-\mu_2)^2}{2\sigma_2^2}} dz + \int_T^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(z-\mu_1)^2}{2\sigma_1^2}} dz \quad (1)$$

where  $w_1$  and  $w_2$  represent the two data classes, of which there are  $N_1$  and  $N_2$  samples, respectively,  $\mu_1$  is the mean of the  $w_1$  training dataset,  $\mu_2$  is the mean of the  $w_2$  training dataset,  $\sigma_1$  is the standard deviation about  $\mu_1$ , and  $\sigma_2$  is the standard deviation about  $\mu_2$ . The point  $T$  is determined by differentiating Eq. (1) and setting equal to zero [31]. Equation (1) can easily be expanded for multi-class classification algorithms. Because supervised learning is utilized in this study, the distribution of the training data is known and can be used to determine the distinguishability power of each feature over the training set. This information is used as a means of determining which features have the best ability to distinguish healthy and damaged classes.

**2.3.2 Fisher's Ratio.** Fisher's ratio is a single value that quantifies the distance between the means of two distributions with regard to the variance of each distribution [20]. Equation (2) presents Fisher's ratio

$$FR = \frac{(m_i - m_j)^2}{\nu_i + \nu_j} \quad (i \neq j) \quad (2)$$

In this equation,  $m_i$  is the mean,  $\nu_j$  is the variance, and subscripts  $i$  and  $j$  denote the two classes (healthy and damaged). Equation (2) can be easily expanded to multiclass classification with minimal effort. Fisher's ratio is a measure for the (linear) discriminating power of some variable. A high score indicates a Fisher ratio in which the means are more separable. Features with a larger ratio are assumed to be more useful in differentiating between classes.

## 2.4 Machine Learning Selection/Development

**2.4.1 Overview.** The data-driven approach to structural health monitoring uses statistical pattern recognition and machine learning to monitor a structure and determine whether a fault is present. A benefit of data-driven methods is that they do not require much advanced knowledge of a structure (i.e., material properties and failure mechanisms) to monitor the health of the structure [33]. Certain steps, such as feature scaling and mean normalization, are taken prior to applying a machine learning algorithm, to give it

the highest chance of success possible. These steps help to normalize the data from each feature and help the algorithm to converge on an optimal solution faster. This minimizes the problem of features with a greater numerical range from dominating those with a much smaller range. Additionally, the initial dataset is broken into three subsets: training, test, and cross-validation. The algorithm is trained using the training set, and then evaluated using the test and cross-validation sets to determine how well the algorithm, once trained, can generalize to new examples.

“Machine learning” refers to developing and training an algorithm to recognize patterns and make predictions, either with or without knowledge of preceding examples. Three general categories of algorithms are used in statistical modeling: group classification, regression analysis, and outlier detection. Determining the best category from which to select an algorithm to use depends on whether supervised or unsupervised learning is desired [25]. Supervised learning (linear/logistic regression and support vector machines) relies on knowledge of preceding examples to predict future results. In unsupervised learning (clustering, neural networks, and recommender systems), patterns are inferred from the data with no knowledge of preceding examples [34]. With proper feature selection and an appropriate algorithm, the presence of damage in a structure can be monitored. This includes damage severity, location, and remaining useful life. For a high-cost structure (such as bridges or space vehicles), inciting damage is not practical. In these cases, unsupervised algorithms are typically used though they are limited to the identification of damage [20,25]. When data are known from both a damaged and undamaged state, supervised learning is preferred as it allows extent and location of damage, as well as the remaining useful life, to be determined [20].

In this study, a set of relatively low-cost subscale composite turbine blades are used, allowing damage to be incited, and supervised learning to be used to observe how the acoustic signature changes with damage. The efficiency and accuracy of selected algorithms are monitored, as is feature sensitivity to environmental and operational conditions. Here, the primary concern is structural damage detection. Future studies will implement results from this research on larger turbine blades where additional information regarding the health state, such as extent and location of damage, as well as remaining useful life will be sought. Logistic regression and support vector machines (both supervised learning algorithms) are discussed in detail in this paper. For both algorithms, the process flow is similar: a training set of  $n$  samples (each consisting of the reduced feature vector as well as a class label (healthy/damaged)) is provided to the algorithm. This information is weighted by an initial set of parameters,  $\theta$ , to develop a hypothesis function. The hypothesis is iteratively updated to find the best  $\theta$  matrix to fit the data. Once “trained,” the algorithm can classify future input into its respective classes as depicted in Fig. 3.

**2.4.2 Logistic Regression.** A binary classification logistic regression-based algorithm is selected to monitor the existence of damage on a turbine blade. Due to the binary nature, the system does not show the extent of damage, only that a threshold has been breached that indicated damage of a predetermined severity

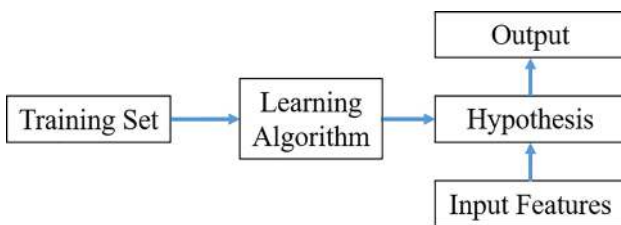


Fig. 3 General process overview for the supervised learning algorithms

is present. Intervention may be a visual inspection, repairing or replacing the blade, or making other changes to the turbine to inhibit further damage. Logistic regression sorts input feature vectors into classes (either binary or multiclass) through learned optimization of a weighting parameter matrix,  $\theta$ . In the current case study, these feature inputs consist of statistical parameters in both time and frequency domain. As stated, acoustic pressure time data are acquired from an external microphone while the structure is internally excited by a known acoustic input. The acquired time data are condensed into a vector of  $n$  features which are compiled into a matrix of  $m$  training examples given in the following equation:

$$\mathbf{x} = \begin{bmatrix} x_1^1 & \dots & x_1^n \\ \vdots & \vdots & \vdots \\ x_m^1 & \dots & x_m^n \end{bmatrix} \quad (3)$$

The feature matrix,  $\mathbf{x}$ , weighted by the set of parameters,  $\theta$ , defines the hypothesis function which attempts to predict the known output classes of the training set. Binary classification yields two possible output values of  $y=0$  (healthy) and  $y=1$  (damaged). Thus, the sigmoid function is used as the hypothesis depicted below

$$\mathbf{h}_\theta(x) = \mathbf{g}(\theta^T x) \rightarrow \mathbf{g}(z) = \frac{1}{1 + e^{-z}} \rightarrow \mathbf{h}_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \quad (4)$$

Figure 4 displays the sigmoid function, plotted as a function of  $z$  (where  $z$  is taken to be all real numbers).

The sigmoid function,  $\mathbf{g}(z)$ , is asymptotic at values of 0 and 1, and greater than 0.5 for any input,  $z$ , greater than 0. In terms of the hypothesis, this means that  $\mathbf{h}_\theta(x) \geq 0.5$  will be generated when  $\theta^T x \geq 0$ . In binary classification, a threshold is set such that any hypothesis value produced above this value corresponds to a predicted class of 1 (damage present) and conversely predicts a value of 0 when  $\theta^T x \leq 0$ . Developing a strong hypothesis involves developing an appropriate parameter set. This is an iterative process, in which the “cost” is computed to determine how well the algorithm classifies the provided training data after each iteration. The output weighting parameters  $\theta$  that generate the lowest cost, based on the initial training set, are then used in the hypothesis to predict future output values. The cost function penalizes the hypothesis heavily each time it predicts a wrong value of  $y$ , and minimally when a correct value is predicted as given by the following equation:

$$\text{cost}(\mathbf{h}_\theta(x), y) = \begin{cases} -\log(\mathbf{h}_\theta(x)) & \text{if } y = 1 \\ -\log(1 - \mathbf{h}_\theta(x)) & \text{if } y = 0 \end{cases} \quad (5)$$

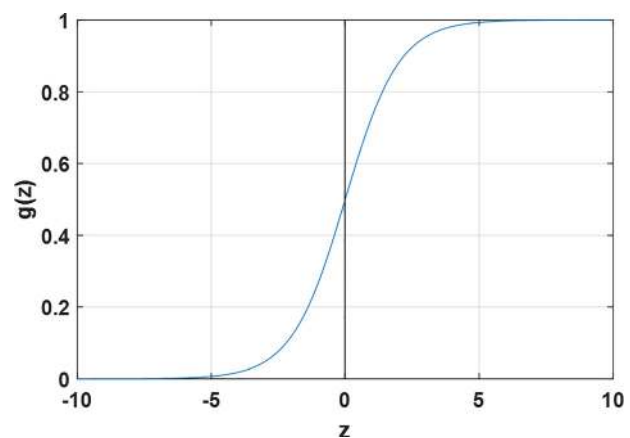


Fig. 4 The sigmoid hypothesis function,  $\mathbf{h}_\theta(x) = \mathbf{g}(z)$

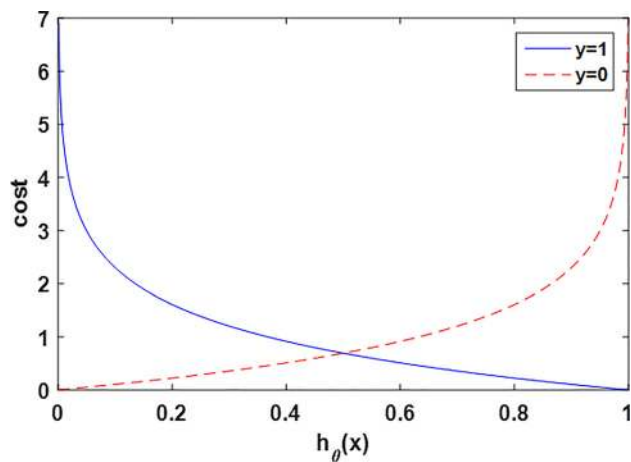


Fig. 5 Logistic regression cost curves for  $y = 1$  and  $y = 0$

Figure 5 depicts the cost function versus the hypothesis, plotted for instances  $y = 1$  (damaged) and  $y = 0$  (healthy).

In Fig. 5, it is illustrated that the cost is high when the hypothesis yields an output contradicting the actual value for  $y$  and vice versa for correct predictions. When calculating the cost, it is important to avoid overfitting in which the algorithm learns the input/output relationship of the training set, but cannot generalize to future examples. This occurs when there are too many features for the number of training examples. Overfitting can be addressed by either reducing the number of features, which is not ideal if all the features are important, or employing regularization. Regularization adds a term,  $\lambda$ , to the cost function, which helps to keep  $\theta$  small essentially “absorbing” the function’s tendency to try and fit every data point [35]. The two-piece cost function in Eq. (6) with the embedded regularization parameter can be summed across all examples in the training set,  $m$

$$J(\theta) = -\left[ \frac{1}{m} \sum_{i=1}^m y^{(i)} \log(\mathbf{h}_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - \mathbf{h}_\theta(x^{(i)})) \right] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2 \quad (6)$$

The *gradient descent* algorithm is used for minimizing the cost function with respect to  $\theta$  [35]. Once the optimal vector  $\theta$  has been determined, the hypothesis function can be plotted representing the boundary at which the algorithm switches between predicting one class or another. The optimized hypothesis now allows future feature sets to be classified. As the number of features and the number of output classifications increase, the decision boundary can become highly complex. Computation time is expensive, particularly in regard to a real-time health monitoring system. Therefore, care is taken to make the iterative solver run as efficiently as possible by applying feature scaling. Features can have ranges that are quite different from each other, resulting in extended computation time to converge. Taking advantage of its superior optimization capabilities, the MATLAB function *fminunc*, which uses Broyden–Fletcher–Goldfarb–Shanno-based optimization, was used [36]. This compensates for range differences and does not require manual selection of a learning rate.

**2.4.3 Support Vector Machines.** Support vector machines (SVMs) are a type of supervised, classification-based machine learning algorithm and have only recently become popular in structural health monitoring [20]. SVMs are robust to very large numbers of variables and a very small sample size, which is beneficial in cases where minimal supervised data are available. By controlling the overall complexity of the model, SVMs are

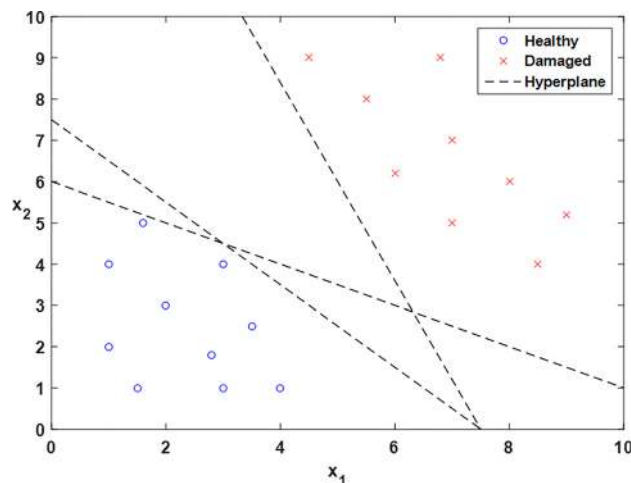


Fig. 6 Multiple hyperplanes fitting sample dataset

typically able to generalize well to new examples and can learn both simple and highly complex classification models [20]. As unified classifiers, SVMs allow many types of discriminant functions to be used with little or no modification to the overall algorithm (i.e., linear, nonlinear, neural networks, and radial-basis) [20]. SVMs also have the ability to produce nonlinear boundaries, unlike logistic regression, by transforming the feature space and placing a linear boundary in this new space. SVMs minimize the influence of points that are well within a class boundary, reducing the amount of data the SVM is actively using to draw a boundary, decreasing solution time [37].

Support vector machines work to identify the optimal separating hyperplane. This is done by finding the hyperplane which maximizes the margin between output classes. Equation (7) presents the formulation for a hyperplane

$$\mathbf{w}^T \mathbf{x} = 0 \quad (7)$$

In Eq. (7),  $\mathbf{w}$  and  $\mathbf{x}$  are vectors, the multiplication of which represents the equation of the hyperplane in  $n$  dimensions. This notation is used because it guarantees that  $\mathbf{w}$  is always normal to the hyperplane, and it is easier and faster notation when hyperplane dimensions exceed two. In two dimensions, the hyperplane is a line which is drawn to differentiate between classes. Figure 6 displays a two-dimensional plot, with healthy and damaged data points linearly separable.

Figure 6 displays the many lines which can be drawn, effectively separating the healthy and damaged classes. Given only this training data, each of the lines drawn is successful in separating classes. However, certain hyperplanes are better than others at generalizing to new examples. Hyperplanes that pass very closely to a training data point are more susceptible to misclassification when new test data are introduced. Hyperplanes that are further from existing training data are more robust to the introduction of new test data. The goal of SVMs is to select the hyperplane which has the maximum margin. The margin is essentially twice the magnitude of the perpendicularly projected distance of the nearest point to the hyperplane. Figure 7 displays the manner of projecting the point distance onto the perpendicular vector of the hyperplane.

Figure 7 displays the vector  $\mathbf{p}$ , which is the projected distance of point A to the perpendicular vector  $\mathbf{w}$  to the hyperplane. Twice this value is indicative of the margin of the particular hyperplane. Figure 8 displays two cases in which the margin of an ill-fitting hyperplane and an optimal hyperplane is shown. It can be seen in Fig. 8(a) that the margin is relatively small. A new data point introduced which falls just outside the existing cluster of healthy data but still far removed from the damaged cluster could easily

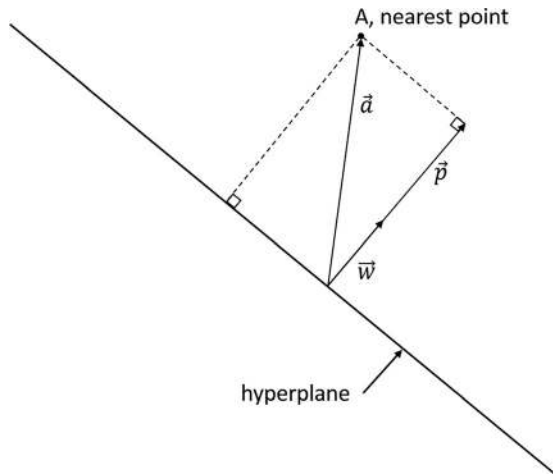


Fig. 7 Illustration of the steps in hyperplane formulation

be misidentified as damage. This is undesirable in that, from a structural health monitoring perspective, detecting damage when none is present can result in added expense and downtime for a structure.

It can be seen in Fig. 8(b) that this hyperplane is much more robust to the case where a new data point falls slightly outside of the existing clusters and is less likely to misidentify a data point that still maintains characteristics of a particular class and falls near to that class.

Similar to logistic regression, the optimal SVM parameters (the best hyperplane) are found by minimizing a cost function which evaluates the performance of the algorithm being trained in comparison to the output of the existing training set. SVMs are convex functions, which imply that any local optimum is also the global optimum. Though the cost function is formulated slightly differently than that of logistic regression, the optimum algorithm parameters are found by minimizing the respective cost function. Figure 9 displays the two part cost function for cases where  $y = 1$  and  $y = 0$ . For comparison, the logistic regression cost function for each case is overlaid in a dotted line.

Combining the terms for  $y = 1$  and  $y = 0$ , the global cost function for SVMs is given in the following equation:

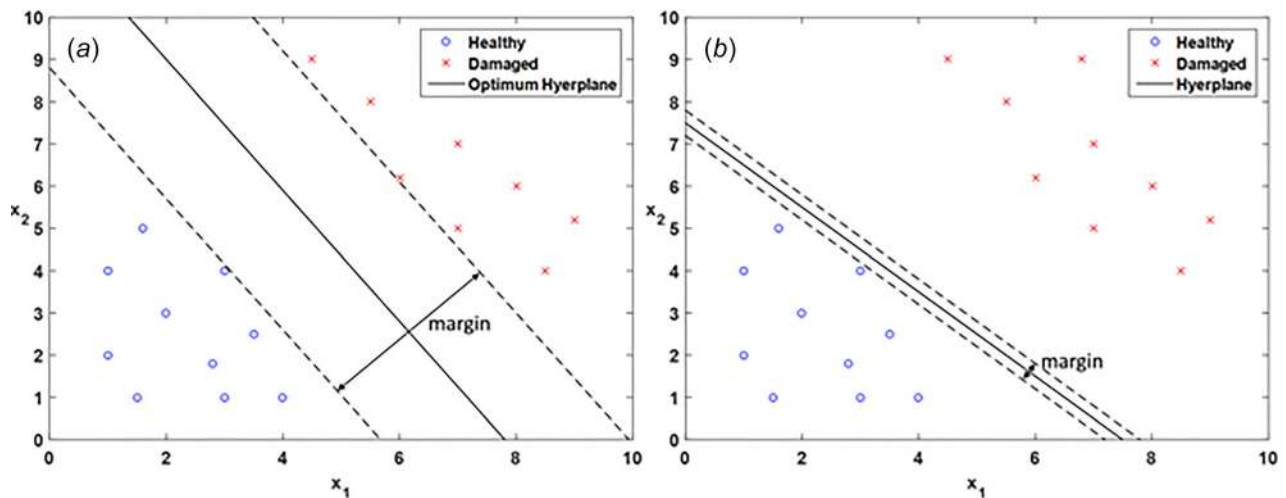


Fig. 8 Ill-fitting and optimal hyperplane comparison

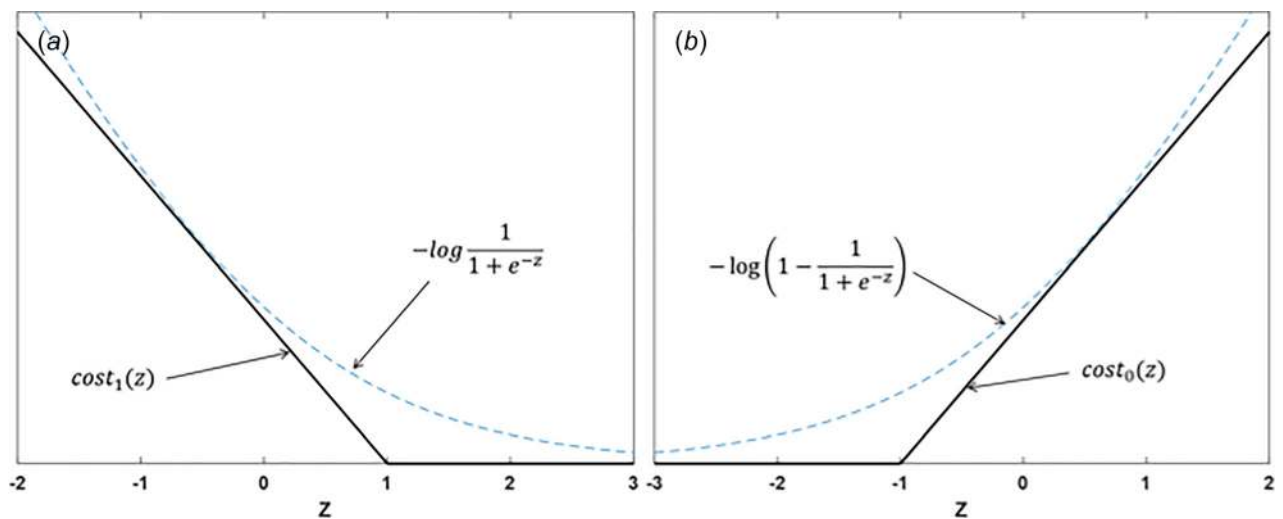


Fig. 9 Cost functions for (a)  $y = 1$  and (b)  $y = 0$

$$\mathbf{J}(\theta) = C \sum_{i=1}^m \left[ \mathbf{y}^{(i)} \cos t_1 (\boldsymbol{\theta}^T \mathbf{x}^{(i)}) + (1 - \mathbf{y}^{(i)}) \cos t_0 (\boldsymbol{\theta}^T \mathbf{x}^{(i)}) \right] + \frac{1}{2} \sum_{j=1}^n \theta_j^2 \quad (8)$$

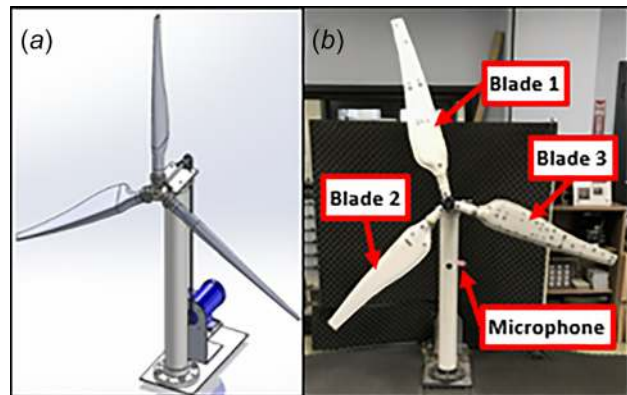
It can be seen in Eq. (8) that the regularization parameter  $\lambda$  is no longer present in the second term of the equation, and that a new parameter  $C$  has been introduced.  $C$  acts against overfitting in the presence of outliers. If  $C$  is very large, the decision boundary will be largely influenced by outliers, whereas conversely, if  $C$  is very small, the algorithm will largely ignore the presence of outliers. For this paper, MATLAB's toolbox is used to evaluate SVMs for SHM applications [36].

**2.5 Limitations.** In a realistic situation, it is unlikely to have access to a sufficiently large amount of labeled data from full-scale wind turbine blades that can be leveraged for data classification using supervised learning. The currently envisioned damage detection system involves a two-stage detection and classification algorithm. The first stage is to do anomaly detection using unsupervised learning (clustering-based) for damage detection. This is based upon the principles of data clustering, i.e., newly collected data having statistical departure from the baseline (healthy) new blades. The baseline data will be collected by the system using the second-stage algorithm to train itself for damage identification (not only detect but also localize and predict the severity). The second stage will allow increasing the labeled database as the data will be labeled right after a new turbine (no blade damage initially) is deployed in the field (healthy baseline data) and during maintenance stages (healthy baseline or damaged, if any). The data collected and processed (using supervised learning) in this study and the future controlled experimental studies conducted by the team are geared toward developing an initial database (data collected from different blades and under different conditions) and develop the second stage of the operational damage detection system. This initial experience will be complemented by data obtained from wind farms and in situ application of the system and is critical for the success of the proposed damage detection system.

The test structure under investigation is a subscale model of an actual wind turbine system that is tested in a controlled laboratory environment. A fully operational wind turbine system will be subjected to various internal and external noise sources including the wind flow over the blades, whining of the gearbox, and other environmentally induced noise. Some noise will be clearly distinct or tonal, such as the gearbox-related part, and can be simply filtered out. Most noise will be complex, such as the flow-induced noise, and will require the use of advanced signal processing algorithms to account for the influences on the data, and subsequently the machine learning algorithm. Further investigation into the characteristics of tonal and nontonal noise contributors will include wind tunnel testing and active detection tests in the field on a real wind turbine system. For the purpose of this paper, the expected noise contributions in the field are ignored with the intent to provide a proof of concept. In response to the success of the investigated machine learning algorithms and active detection concept, further development and testing can be performed to address any noise concerns.

### 3 Test Structure

One of the primary objectives of this study is to select machine learning algorithms that can detect the presence of damage. The capability to detect damage should be unaffected by the type, severity, and location in which the damage can exist. In order to assess whether the selected machine learning algorithms provide sufficient capability to detect damage despite the various circumstances in which damage can exist, a practical and large test



**Fig. 10** (a) Solid model of the subscale turbine and (b) completed subscale turbine prototype

matrix was developed and executed. The results obtained from executing all the tests provide enough data to properly evaluate the machine learning algorithms. A laboratory-scale wind turbine (rendered model in Fig. 10(a)) was designed to test the effectiveness of the selected machine learning algorithms. The completed wind turbine (Fig. 10(b)) has three hollow composite blades, each containing a wireless speaker. The composite blades were manufactured using vacuum infusion, which is the preferred manufacturing method in the wind turbine industry. Each blade consists of two halves (clam-shell design) that are sealed together along the edges with sealant putty and acoustic foam tape. The entire test setup was surrounded by acoustic absorbing material in order to minimize sound reflections. Acoustic pressure time history and the corresponding frequency spectrum were recorded using a PCB 130D20 microphone mounted onto the tower, where the microphone was vibration-isolated from the tower using a foam material.

Each wind turbine blade was labeled from 1 to 3 for bookkeeping purposes. The blades were orientated in the same position during all stationary tests, and blade 1 was always used as the damaged blade. Each blade had the same acoustic excitations during multitone tests, and the same volume (amplitude) was used for all speakers. When testing a specific damage type at a specific location, all tests were performed within the same day starting at the healthy state and increasing in damage from small to large. The temperature was monitored for each measurement and recorded due to the known effects of the environmental conditions on acoustic data. All specified precautions were taken in order to minimize the variability among experiments. The following test matrix shown in Table 1 presents all states in which the blades were tested. A total of 28 different blade states were performed, in which four were tested while the wind turbine hub was rotating. During these tests, the rotor hub's rotational speed was controlled using a direct current motor.

Two damage types were considered in this paper, holes and edge splits. Holes provide a simple (but still realistic) form of damage to understand and implement, which is done with a hand drill and drill bits. Edge splits resemble a practical and common type of damage and are created by removing the sealant putty and acoustic foam tape sealing the two edges together. Holes are repaired by using Bond-O all-purpose putty. Edge splits are repaired by reapplying the sealant putty and acoustic foam tape along the edge where damage was induced. Damage types were individually implemented at the root, midlength, or tip of the blade to cover a range of locations in which damage could occur. A full schematic of each damage type and location on blade 1 is shown in Fig. 11. Each damage was tested at four levels of severity classified as healthy, small, medium, or large in order to assess the sensitivity of the machine learning algorithm. The dimensions of each damage severity are shown in Table 1. Dimensions

**Table 1 Summarized test matrix of all test cases performed**

Operation	Type	Location	Severity
Stationary	Hole	Root	Healthy (0.000 in)
			Small (0.125 in)
			Medium (0.250 in)
			Large (0.375 in)
		Midlength	Healthy (0.000 in)
			Small (0.125 in)
			Medium (0.250 in)
			Large (0.375 in)
		Tip	Healthy (0.000 in)
			Small (0.125 in)
			Medium (0.250 in)
			Large (0.375 in)
Edge split	Root	Healthy (0.000 in)	
		Small (1.000 in)	
		Medium (2.000 in)	
		Large (3.000 in)	
	Midlength	Healthy (0.000 in)	
		Small (1.000 in)	
		Medium (2.000 in)	
		Large (3.000 in)	
	Tip	Healthy (0.000 in)	
		Small (1.000 in)	
		Medium (2.000 in)	
		Large (3.000 in)	
Rotational	Hole	Midlength	Healthy (0.000 in)
			Small (0.125 in)
			Medium (0.250 in)
			Large (0.375 in)

represent the diameter of the holes and length of the edge splits, respectively.

Once one damage type and location had been tested at every severity, the blade was repaired to a "healthy" state as outlined. The repaired blade acts as the new baseline healthy state for the next damage type and location. Therefore, any possible influences from the repairs performed on the blade are neglected when considering a damage type and location. A set of experiments were performed with the wind turbine blade in operation. The blades were set to rotate at a speed of approximately 45 rpm. The damage case considered during the rotating test was an edge split located near the midlength of blade 1.

Seven acoustic excitation types were used as inputs for each test case to assess how well specific frequency tones or ranges can detect damage, as well as how sensitive the frequencies are in the

**Table 2 Acoustic excitation parameters for each test case**

Test	Blade 1 (Hz)	Blade 2 (Hz)	Blade 3 (Hz)
Single low	1000	—	—
Single mid	5000	—	—
Single high	10,000	—	—
Multi-low	1000	900	1100
Multi-mid	5000	4500	5500
Multi-high	10,000	8000	9000
White noise	White noise	—	—

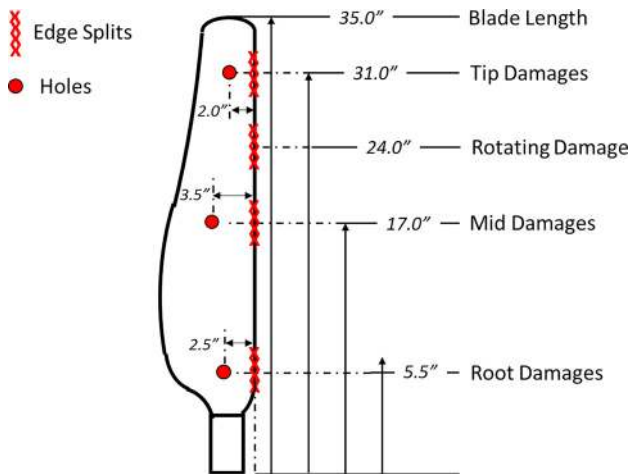
presence of other tones. The specific frequencies and which blade is excited for each test are shown in Table 2. Single-tone harmonic (sinusoidal) tests ranging from 1 to 10 kHz excite only blade 1, the damaged blade, to assess the detectability of damage at low, medium, and high frequencies. Using a single tone asserts that only the damaged blade is in consideration, the simplest form of excitation is used, and all the sound energy is focused at one frequency for ease in distinguishing results. Multitone sinusoidal tests excite each blade at similar (but still different) frequencies to assess how well damage can be detected when multiple tones exist. Additionally, ensonifying each blade with a separate frequency increases the likelihood of identifying which blade has damage based on the frequency that indicates damage. Blade 1 is excited with the same frequencies as was done with the single-tone tests to reduce variability. For each multitone range tested, blades 2 and 3 are excited at frequencies one-tenth less or greater than the frequency of blade 1. White noise tests excite only blade 1 to assess the ability to detect damage with a broadband frequency input. The data from a white noise test can be observed to determine which frequency ranges within the input range are optimal or insignificant for damage detection.

Data were acquired using a National Instruments PXI 64 channel system powered by M+P SMART OFFICE data acquisition software. A complete summary of the data acquisition parameters for all tests is shown in Table 3. Parameters were selected such that all frequencies can be acquired with sufficient detail and enough points exist to appropriately represent the features selected in the machine learning algorithms. A total of 30 blocks were taken for each measurement and saved along with an average. Three iterations were performed for each measurement to give a total of 90 time blocks for each acoustic excitation in each test case. For example, the healthy hole located at the tip of the blade will have 90 time blocks for each acoustic excitation adding up to 630 healthy time blocks. In total, 17,640 time blocks were acquired and saved in which 4410 blocks represent healthy states. This allows for a statistically significant quantity of data to train the machine learning algorithm.

With the relatively large quantity of data acquired, machine learning algorithms' ability to detect the presence of damage can be sufficiently tested under multiple damage types, severities, and locations. Including and excluding datasets from specific tests enable thorough assessments of the algorithms' capability with respect to acoustic excitation types and damage characteristics. For example, some of the acoustic excitation types may be poor indicators when detecting damage for some or all features.

**Table 3 Signal processing parameters for each test**

Variable	Value
Sample rate (Hz)	32,768
Block size (no. points)	131,072
Frequency resolution (Hz)	0.25
Time resolution (s)	$3.05176 \times 10^{-5}$
Sample time (s)	4
Averages	30
Window	None



**Fig. 11 Schematic of blade 1 damage locations**



Therefore, these acoustic excitation types can be excluded when training and testing the machine learning algorithms. A parametric analysis is performed considering various combinations of the data to identify optimal features, excitation types, and conditions that ultimately provide the capability to detect damage from the selected machine learning algorithms.

#### 4 Results and Discussion

A test case is defined as a complete run through of all damage severities of a specific damage type and location for a single acoustic excitation type. For example, a hole-type damage located at the root of the blade acoustically excited with white noise is classified as one test case. It is assumed that if the accuracy of the machine learning algorithms is poor when considering an individual test case, it will also be poor when considering that test case in conjunction with other test cases. Therefore, each individual test case is considered for evaluation first. Once every test case has been evaluated, the test cases that were accurately classified are considered for grouped analysis. The resultant accuracy of the grouped test cases allows proper conclusions to be drawn about the sensitivity and ability of the machine learning algorithms to detect damage when subject to different damage types, severities, and locations. The data were filtered with either a high-pass or band-pass filter depending on the acoustic excitation method. Undesired low-frequency content has a significant influence on the features. Filtering as such allowed the time data to maintain the contributions of the damage on the data while excluding noise from the laboratory environment, equipment, and the wind turbine blade motor. The filtered data are organized into features and normalized. The resultant feature matrix is used in the logistic regression algorithm and SVM algorithms.

**4.1 Evaluation of Machine Learning Algorithms.** As previously mentioned, logistic regression machine learning algorithms were implemented in MATLAB. The data were split into three groups: the first for training the algorithm and the remaining two for evaluating the algorithm's ability to generalize to new examples. Support vector machines are executed directly from the MATLAB's Classification Learner application. Six types of SVMs are considered: linear, quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian. Linear, quadratic, and cubic make a polynomial separation of the respective order between classes. The Gaussian SVMs make distinctions between classes using a Gaussian kernel. The terms fine, medium, and coarse refer to the kernel scale in which the Gaussian kernel is based. The kernel is calculated based on the number of predictors (features). As the kernel decreases in size, the distinction between classes is finer and captures detailed intricacies in the feature relationships. Finer kernels may be required for distinguishing complex feature relationships, but are easily prone to overfitting.

Figure 12 displays the number of instances that each machine learning algorithm was able to accurately classify the presence of damage with respect to each acoustic excitation type when considering each individual test case. The test case was considered acceptable from the classification perspective, if the accuracy was

greater than 98%. A 98% accuracy corresponds to no more than 4 out of the 216 errors considered during training, 1 out of 72 during testing, or 1 out of 72 during cross-validation. High accuracy in the training and cross-validation sets implies that the logistic regression algorithm was able to generalize to the test and cross-validation sets indicating no issues exist with overfitting. A total of seven different damage type and location combinations were tested; thus, a value of seven indicates that the algorithm accurately classified the damage data for all damage types and locations. Values of seven are bolded and highlighted while values greater than five are bolded.

Figure 12 gives insight into two aspects of the designed methodology, namely, the quality of acoustic excitation method used and the quality of the machine learning algorithms tested. In terms of acoustic excitation, the best accuracy was received using the multi-mid acoustic excitation. Second to the multi-mid excitation is the multi-high acoustic excitation, closely followed by the 5000 Hz sinusoid and white noise excitations. In regard to the machine learning algorithms, each managed to perform with significant and comparable accuracy. However, it is clear that the fine Gaussian SVM performed poorly. The most accurate algorithm was the quadratic SVM, which would be expected due to the additional leniency in defining the boundary between the two cases while maintaining a sufficient margin for future generalization.

Based on how well the machine learning algorithms performed for the individual test cases during stationary tests, multiple test cases were selected to be evaluated collectively by the algorithms. For example, it was found that using a 5000 Hz acoustic excitation yielded well-trained algorithms in all damage cases but those located at the root. Therefore, the algorithms were provided all the data from the 5000 Hz test cases excluding the hole and edge split test cases located at the root. It can also be noted that multi-low excitations performed similarly to 5000 Hz excitations, multi-mid excitations worked well with all cases, multi-high excitations could accommodate holes perfectly, and white noise was sufficient to classify edge splits perfectly. Figure 13 shows the combined test cases considered and the resultant accuracies for each model. Accuracy values that are greater than 98% are highlighted again. During the execution of some of the combined tests, it was observed that the features resultant from the first damage state and healthy damage state could not be distinguished. This is not unlikely, as the smallest damage severity considered is quite insignificant. It is plausible that the machine learning algorithm is not able to detect such miniscule changes. Therefore, test cases in which this phenomenon was observed were repeated without the first level of damage. In general, an increase in accuracy was obtained in this case.

Considering each machine learning algorithm, the cubic SVM performed with the greatest accuracy. In fact, the cubic SVM was able to successfully detect the presence of damage with an accuracy greater than 98% for all stationary blade tests when the multi-mid acoustic excitation was used. Linear machine learning algorithms were only able to detect damage groups when a single damage type was considered. Therefore, to include additional damage types, the machine learning algorithms will require a

Excitation	Logistics Regression			Support Vector Machines						Average
	Training	Test	CV	Linear	Quad	Cubic	Fine	Medium	Coarse	
1000 Hz	4	3	4	4	4	4	0	3	3	3.22
5000 Hz	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	0	<b>5</b>	<b>5</b>	4.44
10000 Hz	1	0	1	1	2	1	0	2	1	1.00
Multi-Low	<b>6</b>	4	4	<b>5</b>	<b>6</b>	<b>6</b>	1	<b>6</b>	4	4.67
Multi-Mid	<b>7</b>	0	<b>7</b>	<b>7</b>	<b>7</b>	<b>7</b>	4	<b>7</b>	<b>5</b>	<b>5.67</b>
Multi-High	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	<b>7</b>	<b>7</b>	2	<b>7</b>	<b>3</b>	<b>5.11</b>
Noise	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	<b>5</b>	0	<b>5</b>	<b>5</b>	4.44
Average	4.71	3.14	4.43	4.57	5.14	5.00	1.00	5.00	3.71	

Fig. 12 Instances of accuracies greater than 98%

Excitation	Damages	Logistics Regression			Support Vector Machines						Average
		Training	Test	CV	Linear	Quad	Cubic	Fine	Medium	Coarse	
5000 Hz	All but Root	92.708	93.056	92.361	92.200	98.300	98.200	97.300	98.500	91.900	94.947
Multi-Low	All but Root	81.713	80.903	84.028	83.200	96.000	98.700	97.300	97.200	82.200	89.027
Multi-Mid	All	76.157	74.306	74.537	75.600	97.800	98.300	97.100	95.700	77.000	85.167
	All but D1	76.157	74.306	74.537	75.600	97.800	98.300	97.100	95.700	77.000	85.167
Multi-High	All	73.611	76.157	74.537	76.700	88.500	65.500	94.700	87.200	77.500	79.378
	All but D1	76.955	78.704	75.309	81.900	95.400	96.800	94.800	91.400	75.100	85.152
	All Holes	100.000	97.685	99.537	99.400	100.000	100.000	98.900	100.000	97.800	99.258
	All Holes but D1	100.000	97.531	100.000	100.000	100.000	100.000	99.300	100.000	100.000	99.648
Noise	All but Hole Mid	82.315	85.278	83.056	83.400	90.400	96.300	93.100	87.800	82.200	87.094
	All Edge Splits	98.611	97.685	98.611	98.800	98.600	98.600	95.200	98.700	95.700	97.834
	All Edge Splits but D1	100.000	100.000	100.000	100.000	100.000	100.000	94.800	100.000	100.000	99.422
	Average	87.11	86.87	86.96	87.89	96.62	95.52	96.33	95.65	86.95	

Fig. 13 Combined damage case testing accuracies

Test Cases	Damage Type	Rotating Edge Split						
	Damage Location	Mid-Length						
	Acoustic Excitation	1000 Hz	5000 Hz	10000 Hz	Multi-Low	Multi-Mid	Multi-High	Noise
Logistics Regression	Training	99.54	100.00	100.00	100.00	100.00	100.00	90.74
	Test	100.00	100.00	97.22	100.00	100.00	100.00	88.89
	CV	100.00	100.00	100.00	100.00	100.00	98.61	90.28
	Average	99.85	100.00	99.07	100.00	100.00	99.54	89.97
SVM Metrics	Linear	99.40	100.00	100.00	100.00	100.00	98.90	85.80
	Quadratic	100.00	100.00	99.40	100.00	100.00	99.40	89.20
	Cubic	100.00	100.00	100.00	100.00	100.00	99.40	86.90
	Fine Gaussian	86.70	90.80	86.40	78.30	88.60	93.90	74.70
	Medium Gaussian	100.00	100.00	99.20	100.00	100.00	99.40	78.60
	Coarse Gaussian	99.20	100.00	100.00	100.00	100.00	96.70	75.00

Fig. 14 Machine learning accuracies for rotating tests

projection into a higher feature space to make appropriate decisions about the data.

It is clearly observed that the multi-high-type acoustic excitation is unsurmountable at distinguishing the presence of hole-type damages. Similarly, the white noise acoustic excitation excels when detecting edge split damages. Explanations as to why certain acoustic excitations work better than others are not completely understood at the current state of this work and need to be further studied in the future. However, it is expected that complex interactions exist between the sound pressure and damage geometry that significantly influence the sound field external to the wind turbine blade. Depending on the size and shape of the damage, radiated sound will exhibit different directivities when exiting through the damage. Each blade forms an acoustic cavity that has an associated set of acoustic modes. Some frequencies may produce a larger response than others, and the frequency response of the blades is assumed to highly influence the results. In addition, the sound pressure distribution along the internal cavity of the structure will influence the radiated sound energy. For example, if an acoustic node is located at the root position of the blade, it is expected that little variance will exist in the external acoustic field.

The accuracy results from the rotating blade tests conclude the numerical evaluation of the performance of all the machine learning algorithms. Figure 14 displays the accuracy of all the machine learning algorithms for each acoustic excitation type during the rotating blade tests. For almost all acoustic excitation methods, the algorithms were easily able to make the correct decision

between healthy and damaged blade states. The only acoustic excitation method that was unable to enable appropriate damage detection was white noise. Even though a single damage case was tested, the results are promising when considering the implementation of the system on an operational wind turbine.

It is possible for the feature vector to extend to numerous dimensions. In order to represent the classification ability of the

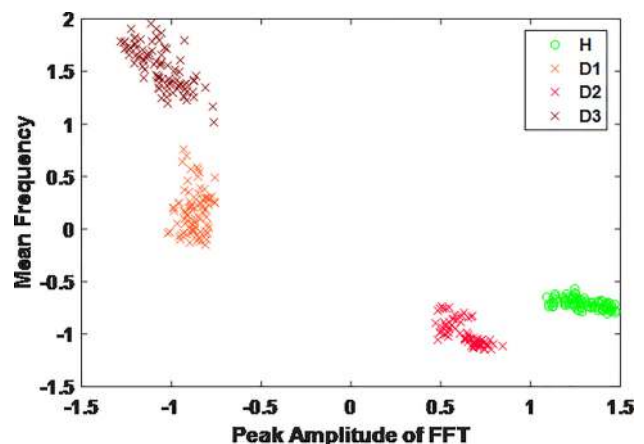
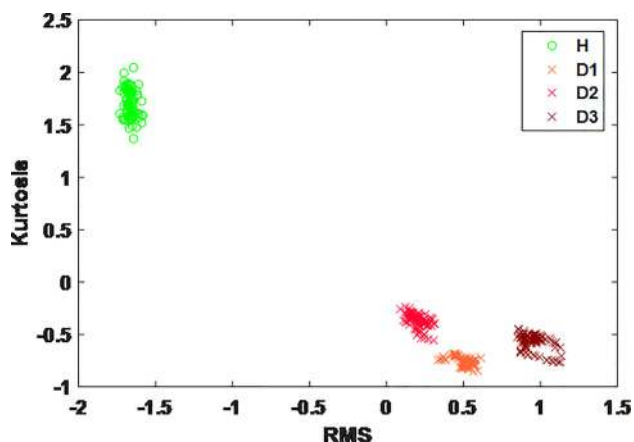
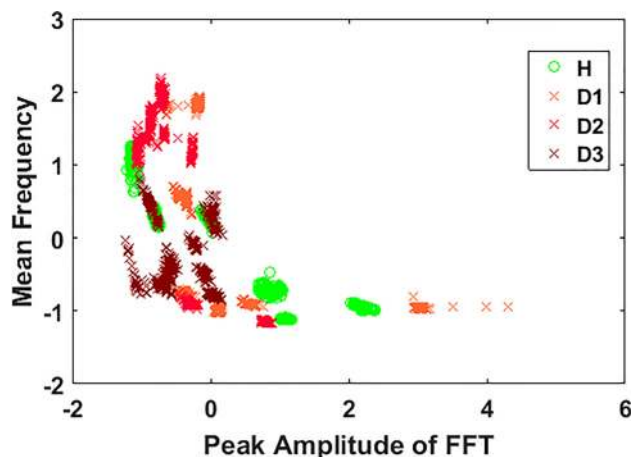


Fig. 15 Comparison of the peak amplitude FFT and mean frequency features for a stationary multi-mid excitation test



**Fig. 16** Comparison of the RMS and kurtosis features for a stationary multi-high excitation test



**Fig. 17** The two-feature plot for all test cases from a stationary multi-mid excitation test

algorithm, two-feature pairs are displayed with respect to various test cases. In all the following classification plots, the healthy data (H) is presented in the form of circles and damaged data as an "x".

Figure 15 displays the data from a multi-mid acoustically excited test considering an edge split damage type located at the midlength of the blade during a stationary test. The two features

compared are the mean frequency and the peak amplitude of the FFT.

It can be seen in Fig. 15 that the healthy and damaged classes are easily separable. The values that correspond to the features change as the damage becomes more severe. Another example of well-distinguished features is shown in Fig. 16 comparing the kurtosis and RMS from a multi-high excitation test considering a hole-type damage located at the tip of the blade.

A comparison of two features from an evaluation when multiple test cases were evaluated at once is shown in Fig. 17 as follows. The results presented consider every damage type using a multi-mid acoustic excitation test.

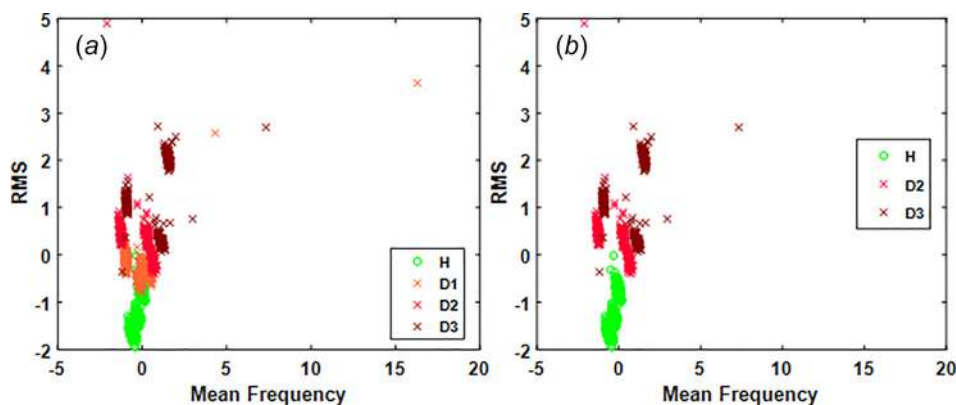
The data are no longer tightly clustered and separable. It is clear that a higher order algorithm is needed to correctly classify the data. This is a direct reflection of what has been observed numerically in Fig. 13, in which the algorithm suffered in correctly classifying the datasets unless a cubic SVM was used.

Figure 18 shows a side-by-side comparison of a two-feature plot in the combined white noise excitation test for all edge split-type damages. Figure 18(a) included all damage severities, and Fig. 18(b) excluded the smallest level of severity.

It is clear that the smallest level of damage is well mixed with the healthy data. When the first level of damage was excluded, the healthy data are easier to distinguish from the damages as shown in the rightmost plot. The augmentation in accuracy is further verified in Fig. 13 where the accuracies of the machine learning algorithm improved to 100% for almost every algorithm.

One of the issues when comparing every damage type is the nonrepeatable location of each healthy dataset across test cases. Figure 19 shows a side-by-side comparison of a two-feature plot for RMS and mean frequency from all hole-type damages using the multi-high excitation parameters. Figure 19(a) includes all the data from each test case. Figure 19(b) only includes the healthy datasets. The legend in Fig. 19(b) indicates which test case the healthy data are from.

There is not a clear distinction or trend between the locations of the healthy clusters. Several reasons could account for the variability. Each test case was performed on a different day; thus, the equipment, temperature, and background noise could be different. In between each test case, the blade is repaired as described in Sec. 3. The influences of each damage or repair on the successively performed test case are unknown but expected to have minimal contributions to the variability. However, the influences from damages and reparations are not expected to be of concern in field tests on a full wind turbine system. Environmental effects are expected to be significant and will need to be further understood in order to account for the influences on the data for the future system.



**Fig. 18** A side-by-side comparison of a feature pair with and without the first level of damage included

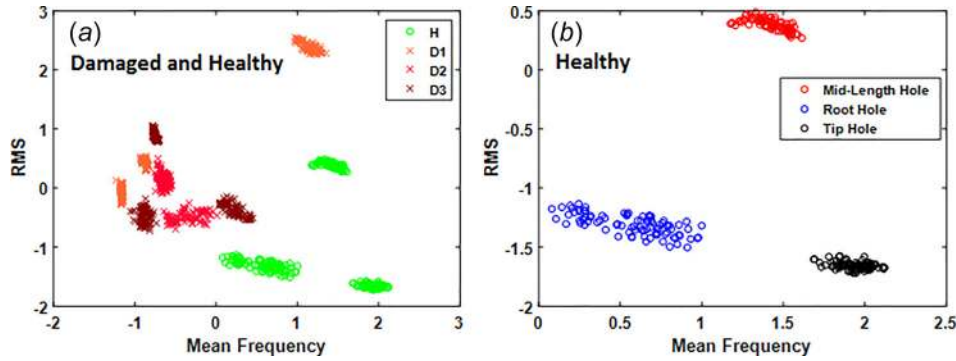


Fig. 19 A side-by-side two-feature plot comparing the variability of healthy data clusters across test cases

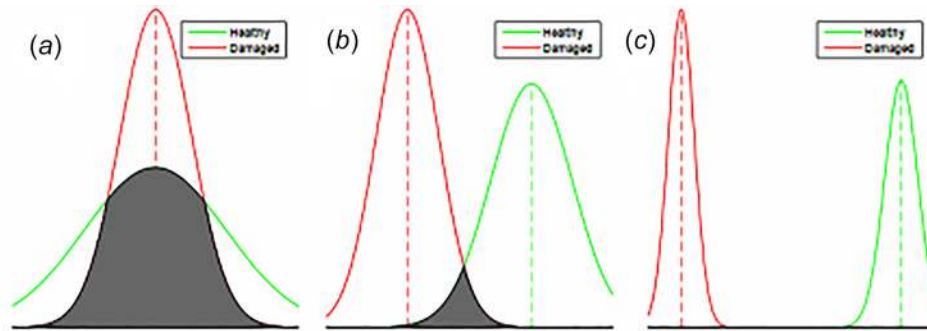


Fig. 20 Distinguishability distribution for (a) case 1, (b) case 2, and (c) case 3

## 4.2 Evaluation of Features

**4.2.1 Distinguishability.** Distinguishability was investigated as a means of developing a quick feature selector. It became apparent that this indicator was much more time intensive than Fisher's ratio. In fact, the distributions of the features were prone to three distinct cases. Figure 20 shows the three distinct cases using the feature distributions from a multi-high acoustic excitation test.

Figure 20(a) shows the first case, in which the distributions intersect twice and are essentially completely overlapped. Figure 20(b) displays a plot where there is only significant overlap between a single tail from each distribution. In this case, the area of overlap can easily be computed. The last case is shown in Fig. 20(c), where there is virtually no overlap, which resulted in an error when finding the point of intersection, as it was below machine precision. Computing the actual area of overlap for all cases would require code which could accurately account for each case, which would require a complex logic. If distinguishability is to be used as a smart feature selector, an approximation is proposed. If there is two intersection points between the two distributions or the distributions are clearly overlapped beyond some threshold value, as in case 1, a value of 1 is assigned, indicating poor distinguishability. If there is only significant overlap between a single tail from each distribution, as in case 2, the distinguishability is determined as the area of overlap. In the final case, of virtually no overlap, a value of 0 is assigned, indicating strong distinguishability. Figure 21 presents the distinguishability results for the multi-high excited test case for a hole damage located at the tip of the blade.

Distinguishability was found to be a more cumbersome method for real-time feature selection than Fisher's ratio, though results were similar in identifying key features.

**4.2.2 Fisher's Ratio.** Fisher's ratio is proposed as a metric for identifying the features from a global list, which are most

successful at discriminating between healthy and damaged classes. In order to appropriately evaluate the Fisher's ratio, only the 49 individual test cases are considered. This way, every acoustic excitation type has the same number of samples. The Fisher's ratios calculated for each individual test case were rescaled to

Feature	Value
RMS	0.000
Kurtosis	0.000
Mean Frequency	0.000
Mean	1.000
Median	1.000
Median Frequency	1.000
RSSQ	0.000
Crest Factor	0.083
Standard Deviation	0.000
Variance	0.000
Skewness	1.000
Peak Amplitude FFT	0.000

Fig. 21 Multi-high tip hole distinguishability results

Feature	>98% Average	All Average	1000 Hz	5000 Hz	10000 Hz	Multi-Low	Multi-Mid	Multi-High	White Noise
RMS	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	0.790	<b>0.982</b>	0.561	<b>0.945</b>	<b>0.994</b>	<b>0.990</b>
Kurtosis	0.255	0.146	0.285	0.110	0.038	0.130	0.535	0.262	0.008
Mean Frequency	0.764	0.672	0.168	0.382	0.422	<b>1.000</b>	0.647	<b>0.949</b>	<b>1.000</b>
Mean	0.000	0.000	0.257	0.078	0.111	0.000	0.000	0.000	0.000
Median	0.095	0.117	0.624	0.063	0.000	0.331	0.039	0.012	0.081
Median Frequency	0.589	0.513	0.070	0.501	0.297	<b>0.856</b>	<b>0.900</b>	0.841	0.025
RSSQ	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	0.790	<b>0.982</b>	0.561	<b>0.945</b>	<b>0.994</b>	<b>0.990</b>
Crest Factor	0.088	0.016	0.000	0.000	0.176	0.056	0.173	0.206	0.002
Standard Deviation	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	0.790	<b>0.982</b>	0.561	<b>0.945</b>	<b>0.994</b>	<b>0.990</b>
Variance	0.847	<b>0.907</b>	<b>0.934</b>	0.573	<b>1.000</b>	0.452	<b>1.000</b>	<b>1.000</b>	0.818
Skewness	0.275	0.215	0.349	0.096	0.007	0.670	0.565	0.015	0.039
Peak FFT	0.726	0.776	<b>0.976</b>	<b>1.000</b>	<b>0.939</b>	0.552	0.850	<b>0.892</b>	0.031

Fig. 22 Fisher's ratio for all tests and excitation types

range between 0 and 1 and averaged with respect to levels of accuracy and acoustic excitation types. The final averages were rescaled once again to range between 0 and 1. Therefore, features which are better at separating the healthy and damaged classes have a score closer to one, and those which are not very capable of distinguishing the classes have a value closer to zero. This allowed general conclusions to be drawn about the overall performance of the features with respect to every test and acoustic excitation type. Figure 22 displays the resultant Fisher's ratios in which perfect scores are bolded and highlighted, and scores above 0.85 are bolded.

First, it can be noted that the RMS, RSSQ, and standard deviation were the most prominent features in detecting damage. This is true for not only all tests but also tests with an accuracy greater than 98%. This is further verified by acknowledging the large number of highlighted cells while scanning across the columns for each feature. The RMS, RSSQ, and standard deviation generally ranked first or second as the best feature for all acoustic excitation types. The peak amplitude of the FFT was far more significant during the single-tone tests as opposed to the multitone tests. This could be because the largest peak in the multitone tests was from a frequency in an undamaged blade. Therefore, as damage was implemented, the peak value would never change. Mean and median frequency features were only significant during multitone tests and the white noise test. It is clear that mean frequency and median frequency would perform poorly during the single-tone tests. The values for the two features should be essentially the same as the frequency of the input tone. However, when a broad frequency range is input in the system, the response could vary significantly when damage is present. Some frequencies may become more dominant in the spectrum as a result of the additional sound energy allowed to transmit through the damage. In all cases, it is noted that the mean, median, kurtosis, skewness, and crest factor almost never performed well in comparison to the other features. The results extracted from the Fisher's ratio are all in agreement with the observed two-feature plot comparisons, as evident from Figs. 15–19. These are preliminary observations on a small series of pilot tests. It is anticipated that the variability due to the blade setup will be minimized when the full-scale blade tests are performed. It should be noted that the goal was not to find a unique feature set for each test group (single versus multitone). Instead, the purpose was to find a means on a per test basis, of isolating and using particularly useful features. It is believed that high variability in test conditions partially hampered the ability to correlate a particular ideal feature set with a particular test condition (single versus multitone tests).

## 5 Conclusions

In this study, the laboratory-scale acoustic tests were performed to understand the damage classification abilities of two supervised learning algorithms, logistic regression and support vector

machines, under different test conditions. Preliminary results indicate normalized Fisher's ratio can be used to determine feature subsets for analysis. Both logistic regression and SVMs were able to classify features into classes with minimal overfitting. Cubic SVMs were required to generalize to all damage types, locations, and severities, but had an accuracy of over 98%. The type of acoustic excitation was studied, and the trends were identified in regard to which excitation method was best for specific types and locations of damage. It was found that white noise was best for detecting edge splits, and multi-high excitations were best for detecting holes. The multi-mid excitation performed well in all cases. The ability to use different kernels allows for nonlinear decision boundaries for the SVM, which is desirable when multiple damage types and locations are considered. When variability is minimized, acoustic damage detection using active monitoring and supervised machine learning appears feasible given this initial study. Future studies will include other means of excitation (chirp and burst random) as well as be on full-scale blades with potential to incorporate blind source separation for determination of which blade is in possession of the fault. The effects of temperature and the environment on acoustic measures will be pursued as well as a method to account for environmental fluctuations. Additional features will be introduced in the machine learning algorithms. Algorithms will be extended to provide the capability of classifying more than damage presence, but the type, location, and severity of damage.

## Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 1538100 (Acoustics-based Structural Health Monitoring of Closed-Cavities and Its Application to Wind Turbine Blades). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## References

- [1] AWEA, 2015, "U.S. Wind Industry Market Reports," American Wind Energy Association, Washington, DC, accessed June 9, 2017, <http://www.awea.org/2015-market-reports>
- [2] DOE, 2015, "Enabling Wind Power Nationwide," U.S. Department of Energy, Washington, DC, Technical Report No. DOE/EE-12187068.
- [3] Lantz, E., Wiser, R., and Hand, M., 2012, "The Past and the Future Cost of Wind Energy," National Renewable Energy Laboratory, Golden, CO, Report No. NREL/CP-6A20-54526.
- [4] Regan, T., Canturk, R., Slavkovsky, E., Niezrecki, C., and Inalpolat, M., "Wind Turbine Blade Damage Detection Using Various Machine Learning Algorithms," ASME Paper No. DETC2016-59686.
- [5] Arora, V., Wijnant, Y. H., and de Boer, A., 2014, "Acoustic-Based Damage Detection Method," *Appl. Acoust.*, **80**, pp. 23–27.
- [6] Yang, W., 2013, "Testing and Condition Monitoring of Composite Wind Turbine Blades," *Recent Advances in Composite Materials for Wind Turbine Blade*, World Academic Publishing, Hong Kong, pp. 147–169.

- [7] Hyers, R. W., McGowan, J. G., Sullivan, K. L., Manwell, J. F., and Syrett, B. C., 2006, "Condition Monitoring and Prognosis of Utility Scale Wind Turbines," *Energy Mater.*, **1**(3), pp. 187–203.
- [8] Kahn-Jetter, Z. L., and Chu, T. C., 1990, "Three-Dimensional Displacement Measurements Using Digital Image Correlation and Photogrammetric Analysis," *Exp. Mech.*, **30**(1), pp. 10–16.
- [9] Ozbek, M., Rixen, D. J., Erne, O., and Sanow, G., 2010, "Feasibility of Monitoring Large Wind Turbines Using Photogrammetry," *Energy*, **35**(12), pp. 4802–4811.
- [10] Niezrecki, C., Avitable, P., Chen, J., Sherwood, J., Lundstrom, T., LeBlanc, B., Hughes, S., Desmond, M., Beattie, A., Rumsey, M., Klute, S. M., Pedrazzani, R., Werlink, R., and Newman, J., 2014, "Inspection and Monitoring of Wind Turbine Blade-Embedded Wave Defects During Fatigue Testing," *Struct. Health Monit.*, **13**(6), pp. 1–15.
- [11] Tipperman, J., and Lanza di Scalea, F., 2014, "Experiments on a Wind Turbine Blade Testing an Indication for Damage Using the Causal and Anti-Causal Green's Function Reconstructed From a Diffuse Field," *Proc. SPIE*, **9064**, p. 90641-1.
- [12] Boukabache, H., Escriba, C., Zedek, S., and Fourniols, J., 2013, "Wavelet Decomposition Based Diagnostic for Structural Health Monitoring on Metallic Aircrafts: Case of Crack Triangulation and Corrosion Detection," *Int. J. Progress Health Manage.*, **4**(3), pp. 1–9.
- [13] Aizawa, K., and Niezrecki, C., 2014, "Wind Turbine Blade Health Monitoring Using Acoustic Beamforming Techniques," *J. Acoust. Soc. Am.*, **135**(4), pp. 2392–2393.
- [14] Aizawa, K., Poozesh, P., Niezrecki, C., Baqersad, J., Inalpolat, M., and Heilmann, G., 2015, "An Acoustic-Array Based Structural Health Monitoring Technique for Wind Turbine Blades," *Proc. SPIE*, **9437**, p. 94371P.
- [15] Fazenda, B. M., 2011, "Acoustic Based Condition Monitoring of Turbine Blades," 18th International Congress on Sound and Vibration (ICSV), Rio de Janeiro, Brazil, July 10–14, pp. 1–8.
- [16] Fazenda, B. M., and Comboni, D., 2012, "Acoustic Condition Monitoring of Wind Turbines: Tip Faults," The Ninth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies (CM and MFPT), London, June 12–14, pp. 1–15.
- [17] Stearman, R. O., Schulz, G. H., and Rohre, S. M., 1997, "Aircraft Damage Detection From Acoustic and Noise Impressed Signals Found by a Cockpit Voice Recorder," *J. Acoust. Soc. Am.*, **101**(5), p. 3085.
- [18] Lam, H. F., Ng, C. T., Lee, Y. Y., and Sun, H. Y., 2009, "System Identification of an Enclosure With Leakages Using a Probabilistic Approach," *J. Sound Vib.*, **322**(4–5), pp. 756–771.
- [19] Nair, K. K., and Kiremidjian, A. S., 2006, "Time Series Based Structural Damage Detection Algorithm Using Gaussian Mixtures Modeling," *ASME J. Dyn. Syst. Meas. Control*, **129**(3), pp. 285–293.
- [20] Sohn, H., Farrar, C. R., Hemez, F. M., Shunk, D. D., Stinemat, D. W., and Nadler, B. R., 2003, "A Review of Structural Health Monitoring Literature From 1996–2001," Los Alamos National Laboratory, Los Alamos, NM, Report No. LA-13976-MS.
- [21] Dervilis, N., Shi, H., Worden, K., and Cross, E. J., 2016, "Exploring Environmental and Operational Variations in SHM Data Using Heteroscedastic Gaussian Processes," *Dyn. Civ. Struct.*, **2**, pp. 145–153.
- [22] Krause, T., Preihs, S., and Ostermann, J., 2014, "Detection of Impulse-Like Airborne Sound for Damage Identification in Rotor Blades of Wind Turbines," Seventh European Workshop on Structural Health Monitoring (EWSHM), Nantes, France, July 8–11, pp. 488–495.
- [23] Edwards, H., Neal, K., Reilly, J., and Van Buren, K., 2016, "Making Structural Condition Diagnostics Robust to Environmental Variability," *Dynamics of Civil Structures* (Vol. 2, Conference Proceedings of the Society for Experimental Mechanics Series), Springer, Cham, Switzerland, pp. 117–130.
- [24] Loh, C., and Chan, C., 2015, "Damage Assessment of Steel Structures Using Multi-Autoregressive Model," *Dyn. Civ. Struct.*, **2**, pp. 1–8.
- [25] Figueiredo, E., Park, G., Farrar, C. R., Worden, K., and Figueiras, J., 2010, "Machine Learning Algorithms for Damage Detection Under Operational and Environmental Variability," *Struct. Health Monit.*, **10**(6), pp. 559–572.
- [26] Nick, W., Shelton, J., Asamene, K., and Esterline, A., 2015, "A Study of Supervised Machine Learning Techniques for Structural Health Monitoring," 26th Modern AI and Cognitive Science Conference (MAICS), Greensboro, NC, Apr. 25–26, Vol. 1353, Paper No. 36.
- [27] Worden, K., and Manson, G., 2007, "The Application of Machine Learning to Structural Health Monitoring," *Philos. Trans. R. Soc., A*, **365**(1851), pp. 515–537.
- [28] Niezrecki, C., and Inalpolat, M., 2015, "Structural Health Monitoring of Wind Turbine Blades Using Wireless Acoustic Sensing," University of Massachusetts, Boston, MA, U.S. Patent No. PCT/US2014/062329.
- [29] Canturk, R., and Inalpolat, M., 2016, "Development of an Acoustic Sensing Based SHM Technique for Wind Turbine Blades," International Modal Analysis Conference (IMAC0XXXIV), Orlando, FL, Jan. 25–28, pp. 95–104.
- [30] Canturk, R., and Inalpolat, M., 2015, "A Computational Acoustic Interrogation of Wind Turbine Blades With Damage," *Comsol Conference*, Boston, MA, Oct. 7–9, pp. 1–6.
- [31] Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., and Wu, B., 2006, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*, Wiley, Hoboken, NJ.
- [32] Farrar, C. R., and Worden, K., 2007, "An Introduction to Structural Health Monitoring," *Philos. Trans. R. Soc., A*, **365**(1851), pp. 303–315.
- [33] Cheng, S., Azarian, M. H., and Pecht, M. G., 2010, "Sensor Systems for Prognostics and Health Management," *Sensors*, **10**(6), pp. 5774–5797.
- [34] Farrar, C. R., and Worden, K., 2013, *Structural Health Monitoring: A Machine Learning Perspective*, Wiley, Hoboken, NJ.
- [35] Le, Q. V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., and Ng, A., 2011, "On Optimization Methods for Deep Learning," 28th International Conference on Machine Learning (ICML), Bellevue, WA, June 28–July 2, pp. 1–8.
- [36] MathWorks, 2014, "MATLAB R2014," MathWorks Inc., Natick, MA.
- [37] Hastie, T., Tibshirani, R., and Friedman, J., 2009, *The Elements of Statistical Learning*, Springer-Verlag, New York.