Guided-Wave-Based Condition Assessment of In-Situ Timber Utility Poles using Machine Learning Algorithms

Ulrike Dackermann[1*], Bradley Skinner[2] and Jianchun Li[1]

[1] Centre for Built Infrastructure Research, School of Civil and Environmental Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney, P.O. Box 123, Broadway, NSW 2007, Australia

[2] ARC Centre of Excellence for Autonomous Systems, School of Electrical, Mechanical and Mechatronic Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, P.O. Box 123, Broadway, NSW 2007, Australia

[*] Corresponding author: E-mail: Ulrike.Dackermann@uts.edu.au; Tel.: +61 2 9514 9064; Fax: +61 2 9514 2633

ABSTRACT

This paper presents a machine-learning-based approach for the structural health monitoring (SHM) of in-situ timber utility poles based on guided wave (GW) propagation. The proposed non-destructive testing method combines a new multi-sensor testing system with advanced statistical signal processing techniques and state-of-the-art machine learning algorithms for the condition assessment of timber utility poles. Currently used pole inspection techniques have critical limitations including the inability to assess the underground section. GW methods, on the other hand, are techniques potentially capable of evaluating non-accessible areas and of detecting internal damage. However, due to the lack of solid understanding on the GW propagation in timber poles, most methods fail to fully interpret wave patterns from field measurements. The proposed method utilises an innovative multi-sensor testing system that captures wave signals along a sensor array and it applies machine learning algorithms to evaluate the soundness of a pole. To validate the new method, it was tested on eight in-situ timber poles. After the testing, the poles were dismembered to determine their actual health states. Various state-of-the-art machine learning algorithms with advanced data pre-processing were applied to classify the poles based on the wave measurements. It was found that using a support vector machine classifier, with the GW signals transformed into autoregressive coefficients, achieved a very promising maximum classification accuracy of 95.7±3.1% using 10-fold cross validation on multiple training and testing instances. Using leave-one-out cross validation, a classification accuracy of 93.3±6.0% for bending wave and 85.7±10.8% for longitudinal wave excitation was achieved.

KEYWORDS

Structural health monitoring, non-destructive testing, timber poles, machine learning, support vector machine, guided waves

INTRODUCTION

For power distribution and communication networks, utility poles form an essential part of a network's infrastructure. Ever since the telegraph was invented, utility poles have been made of timber, and until today, in many countries around the world, the majority of utility poles are still made of wood. As such, in the United States more than 98% and in Australia more than 80% of all utility poles are made of timber, with estimates placing the total number of timber utility poles to around 130 million in the United States and to more than 5 million in Australia (1, 2). Over their design life, timber poles often experience deterioration and decay due to fungus or termite attack, which in most cases are not visible and often below the ground. To avoid failure of utility poles and to ensure the reliability and safety of power distribution and communication networks, utility poles are regularly inspected for maintenance and asset management. Currently, the most commonly used techniques employed by asset management industries are visual inspection, sounding and core drilling (3). Visual inspection is undoubtedly one of the oldest assessment methods used in practice but is limited to accessible areas and surface damage, and like sounding, its reliability and accuracy is highly depended on the experience of the operator. Core drilling is a semi-destructive method that gives only localised information on the tested drilling path. All three methods are based on not measurable parameters and depend on subjective interpretation of information. Neither of the methods is capable of assessing the underground section of a pole, which is indeed the most critical and vulnerable section. These limitations seriously jeopardize the maintenance and asset management and may lead to serious consequences due to undetected faults, as well as unnecessary pole replacements resulting from conservative maintenance approaches. For example, in the Eastern States of Australia, about 300,000 electricity poles are replaced every year despite the fact that up to 80% of the replaced poles are still in a very good serviceable condition, causing a large waste of money as well as natural resources (4). Research has shown that while the current perceived life expectancy of timber poles is approximately 35 years, the average service life can be extended to 75 or more years where "appropriate" inspection and maintenance programs are performed (5).

Limitations of current SHM techniques led to the development of NDT GW-based methods for the condition assessment of pile structures. These methods are potentially capable of detecting internal damage and of evaluating the soundness condition of non-accessible areas such as embedded sections of piles and poles based on the propagation of guided waves. Various types of GW-based methods have been developed such as the sonic echo (SE) method (6-8), the impulse response (IR) method (9), the bending wave (BW) method (10, 11) and the ultraseismic (US) method (12). In the testing, an impact force is generated and the response from the pile structure is recorded by a sensor placed on the pile head. By analysing the reflective wave signals, predictions on the soundness condition of the pile including any damage of the embedded section and length estimations can be made. While GW methods have been used for many years for different types of structures including poles and different materials, their results are still inconsistent due to many issues associated with the complexity of GW propagation including complex wave reflection, attenuation and transformation. The application of GW-based SHM to timber utility poles is in particular very challenging due to the lack of solid understanding of GW propagation in timber pole structures, especially with

the effect of soil embedment coupled with unknown soil and pole conditions below ground line (such as deterioration, rot, termite attack and fungi decay). A major challenge is related to the complexity of the timber material with anisotropic and non-homogeneous characteristics and many uncertainties and variations on material properties (13). As such, the material characteristics of timber can be affected heavily by environmental factors such as temperature and moisture changes, and natural defects, deteriorations and fungi/termite damage cause further complications. As a result, current GW-based NDT methods often fail to fully interpret wave patterns and to produce accurate and reliable condition assessment for timber poles, which is vital for the utility pole management industry.

Because of the stated issues encountered with current GW methods, this paper proposes a new approach using state-of-the-art machine learning classifiers in combination with statistical signal transformation for the non-destructive condition assessment of timber utility poles. The new method presents a solution to currently faced uncertainty issues and provides a reliable testing method that is able to identify faulty poles, which is of crucial importance to pole asset managers. In the proposed method, statistical signal analysis models and advanced machine learning and classification algorithms are applied for feature extraction, pattern recognition and classification of wave signals for the soundness evaluation of in-situ timber utility poles by distinguishing sound from severely unsound structures. A new innovative multi-sensor testing approach based on SE/IR and BW testing is adapted to multiple wave response signals from in-situ timber utility poles. The method is validated on eight in-situ timber poles, which were scheduled for decommissioning. Before decommissioning, the poles were tested in the field using the new GW-based testing method. After testing, the poles were dismembered to determine their actual health states. Based on that autopsy, the poles were classified as either healthy or faulty poles, depending on their individual deterioration state, forming a ground-truth for supervised classification. The results of the proposed machine learning approach for the GW-based condition assessment of timber utility poles show that this technique is capable of overcoming issues encountered by traditional inspection methods and of delivering accurate and robust soundness evaluation results.

BACKGROUND AND METHODOLOGY

The proposed NDT method is based on GW testing with longitudinal and bending wave excitation, and with an array of multiple sensors used to capture the GW response of a timber pole structure. Damage features in the GW signals are extracted using advanced statistical signal transformation based on AR models, and machine learning/classification algorithms are used to determine the health condition of the pole structure by mapping damage features to soundness states.

NDT based on GW propagation

In traditional GW testing for pile-like structures, a stress wave is induced to the structure by applying an impact or impulse to the surface of the structure whereby a sudden pressure or deformation is generated. The disturbance propagates through the structure and is reflected back from changes in stiffness, cross-sectional area and density. The propagation behaviour of the GW is a function of the modulus of elasticity, the density, the Poisson's ratio, and the

geometry conditions of the structure (14). As damage and deterioration changes the structure's properties, the wave propagation behaviour is altered, resulting, for example, in early wave reflection, reduced wave velocity, increased wave attenuation and wave mode conversion. By analysing GW signals through identification of wave velocities, wave reflections and resonant frequency peaks, traditional GW methods, such as the SE/IR method and BW method, aim to detect damage and to determine the dimensions of the structure (e.g. the underground length of a pole structure). The schematic principle of the SE/IR method and the BW method is depicted in Figure 1. For the SE/IR method, the impact is induced from the top of the structure in the longitudinal direction (generating longitudinal compression waves) and wave reflection measurements are recorded by a sensor placed on the top of the structure adjacent to the impact location. Wave signals are analysed in the time domain and the frequency domain, respectively. Details of the SE/IR method can be found in (6-9). For BW testing, a transversal impact is applied to the pile/pole structure generating flexural/bending waves, and wave signals are measured by sensors located on the side of the structure. Thereby, the method is applicable for cases where the top of the structure is obscured such as bridge piles, foundation columns or utility poles. Since bending waves are highly dispersive in nature, dispersive analysis is required in which wave data is extracted from a selected group of frequencies. Details of the BW method can be found in (15-17).

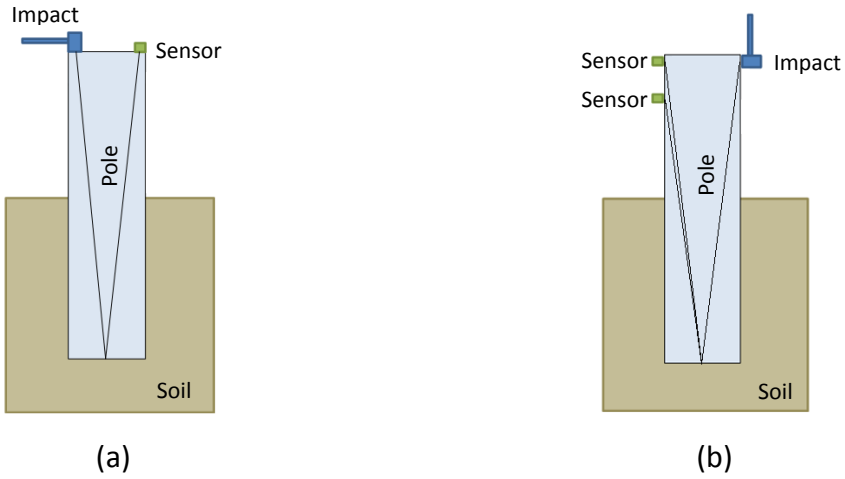


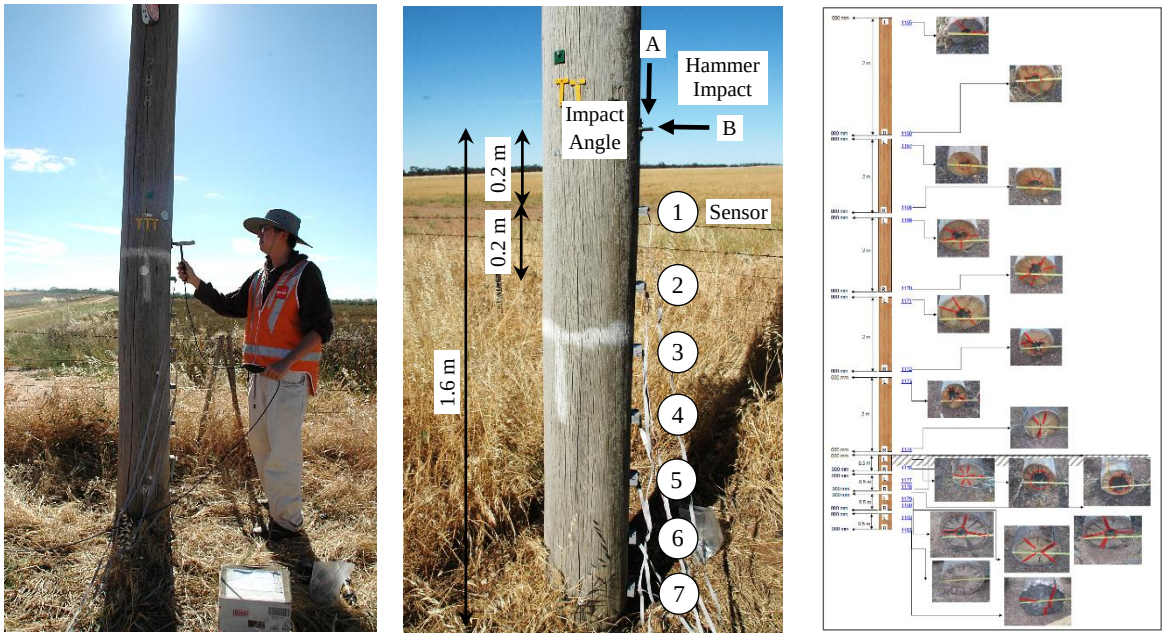
Figure 1. Schematic principle of (a) SE/IR method and (b) BW method.

For the GW testing of pole structures, the generation of GWs from the top of a pole is neither feasible not practical due to their height and the presence of live electricity and telecommunication lines. In the presented testing method, GWs is induced by impacting a pole from its side at a reachable height above ground level in either the transversal direction (analogous to the BW method) generating bending waves, or in the longitudinal direction with an angle (analogous to the SE/IR method) generating primarily longitudinal waves but also bending waves due to the eccentric impact. Because of the impact location, both, up-travel and down-travel waves are generated for each wave type. Both generated types of GWs are low frequency broadband waves that result in multiple wave modes propagating in the pole. To measure the propagating GWs, seven evenly spaced sensors are used, which are placed in a vertical line between the impact location and the soil level. The recording of wave reflection signals from multiple sensors allows the measurement of the waves at different

wave paths leading to a more comprehensive capture of wave features and a more robust condition assessment. For both types of testing (based on the SE/IR method and the BW method), damage patterns inherent in the captured wave signals are extracted using advanced signal processing, and machine learning algorithms are used to identify recurring damage features and to evaluate the soundness of the tested timber utility poles.

Field testing

The proposed method was validated on eight in-situ timber utility poles that were scheduled for decommissioning, and which were tested using a new innovative testing procedure. After the testing, the poles were dismembered to determine their actual health states. Based on that autopsy, the poles were classified as either healthy or faulty poles, depending on their individual deterioration state.



(a) (b) (c)

Figure 2. Field testing and autopsy of in-situ timber utility poles. (a) Execution of field testing, (b) testing set-up and (c) autopsy results of tested and sectioned pole.

The testing execution and set-up is depicted in Figure 2 (a) and (b). An impact hammer was used to generate the GWs and seven accelerometers captured the wave response signals. The impact hammer used was a PCB model HP 086C05 of sensitivity 0.24 mV/N. The hammer impact was induced at a height of 1.6 m in either transverse direction to generate bending waves, or in longitudinal direction (using an impact angle) to generate primarily longitudinal waves. The pole's responses were measured by seven piezoresistive accelerometers that were mounted in a line 0.2 m off ground with spacing's of 0.2 m between the sensors as depicted in Figure 2 (b). The sensors were low cost dual-axis accelerometers of model ADXL320 having a frequency bandwidth of 0.5 Hz to 2.5 kHz and a sensitivity range of 154 to 194 mV/g. These piezoresistive accelerometers were purchased in the form of a circuit board and encased into a specially designed housing. The data acquisition system employed was a mid-range 8 channel system with 12-bit 4M sample/sec per channel model NI PCI-6133. The

signal acquisition and analysis was executed with the National Instrument software Labview. For each test, the sampling rate was set to 1 MHz with a testing duration of 0.5 s, thus capturing 500,000 data points per test. A pre-trigger delay was set to ensure the recoding of the entire impact excitation.

Two types of testing were performed, longitudinal and transverse testing. For longitudinal testing, the impact was executed in the longitudinal direction with the aid of the impact angle and the accelerometers were set to capture acceleration in the vertical direction. For transverse testing, the impact was induced perpendicular to the pole as shown in Figure 2 (a) and the accelerometers measured in the horizontal direction. For each testing type, five tests were performed, i.e. for each pole five hammer strikes were executed in longitudinal direction and five in transverse direction. As seven accelerometer measurements were captured for each testing, a total of 70 measurement signals were recorded for each pole (7 sensors × 5 hammer hits × 2 types of hammer hits).

After the testing, the poles were disconnected from the electricity lines and removed from the soil. To determine the individual health states, each pole was cut into multiple small sections along the cross-section using a chain saw. The exposed cross-sectional areas were photographed and analysed, and according to the seriousness of the found damage and deterioration, the poles were classified as either healthy or faulty. A healthy pole was defined as having only minor to medium damage. A faulty pole was defined as having medium to severe damage with an estimated minimum service life of less than five years. An example of the autopsy of a faulty pole is shown in Figure 2 (c).

Time domain signal segmentation

The measurement data acquired from the five impacts (hammer hits) for both the bending wave (BW) and longitudinal wave (LW) experiments are treated as separate experimental instances that are independent from other instances or impacts. The BW signals illustrated in Figure 3 show the recorded raw data from the seven separate sensor channels for time duration of 0.5 s. After data acquisitioning, the raw data was segmented by removing the pre-trigger data (the first 1300 samples (0.0013 s)) and the steady state data (the last 0.25 s). This effectively reduced the total time-domain stress wave data to a segment size of 248700 samples (windowed: 0.0013 – 0.2500 s). To provide comparative results, the data was also segmented by only removing the pre-trigger data but not the steady state data. Thereby, a segment size of 498700 samples was achieved (windowed: 0.0013 – 0.500 s).

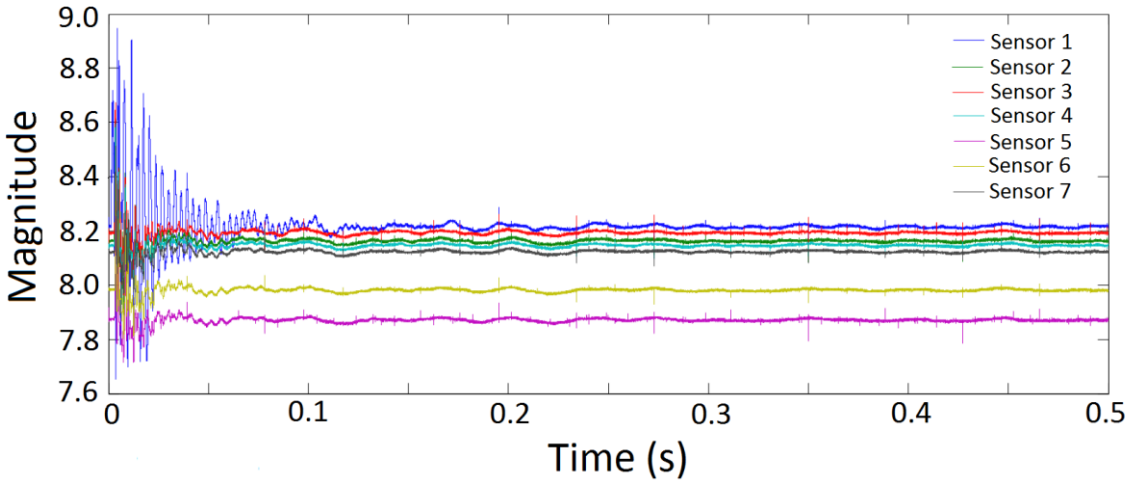


Figure 3. Captured raw BW signals from the seven sensors for a single impact of pole P1.

Parametric methods for stress-wave signal transformation

To extract major features from the recorded stress wave signals (including damage patterns), parametric signal transformation was employed. Signal transformation techniques which estimate the power spectral density (PSD) from a signal directly, such as the periodogram and Welch's method, are commonly known as nonparametric methods. Alternatively, parametric methods estimate the PSD of a signal by assuming it to be the output of a linear system driven by white noise (18). Typical examples of parametric methods include the Yule-Walker, Burg, Covariance and Modified Covariance autoregressive (AR) methods. These AR methods use regression to estimate the PSD by estimating the parameters (coefficients) of the linear system that theoretically "generates" the signal. Parametric methods tend to produce better results than classical nonparametric methods when the data length of the available signal is relatively short (18), which can be modelled as short (quasi)-stationary sequences.

In this study, we employed a parametric representation using AR parameter estimation algorithms to transform the time-domain segmented GW data into a number of real-valued variables. Parametric methods assume that a description of the segmented stress wave signal can be devised from a time-series model of a random process. As such, parametric methods can model fixed segments of stress wave data as the output of a linear filter of order p driven by a Gaussian white noise sequence with zero-mean (18). The output for such a filter is a p^{th} order AR process or maximum entropy method (MEM), given by:

$$x(n) = -\sum_{k=1}^{p} a(k)x(n-k) + u(n), \quad (1)$$

where $x(n)$ is the stationary time-series output sequence that models the fixed segment of stress wave signal data, $a(k)$ are the AR coefficients and $u(n)$ is a Gaussian white noise input driving sequence.

The AR model was used to extract damage-sensitive features, because the underlying linear stationary assumption makes it possible to detect the presence of nonlinearities in the time domain data. The Burg method was used exclusively as the AR parameter estimation algorithm in this study. The Burg method operates on a fixed segment of time samples to

recursively yield a p^{th} order AR model of parameter estimates $a(k)$. The chosen parametric method, which is based on autoregression, transforms the GW input signal from its original time domain representation into a different representation based exclusively on the computed autoregressive coefficients themselves. That is, we directly represent the segmented GW signal using the scalar AR coefficients $a(k)$. As example, Figure 4 shows the resulting AR coefficients computed using the parametric algorithm with an order of p=10 for a segmented BW signal (0.0013 – 0.2500 s) from the seven sensors for a single impact of timber pole P1. Using this approach, we computed the AR coefficients for each timber utility pole (P1 – P8) for all sensors (S1 – S7) according to our desired segmentation of the GW time domain signal.

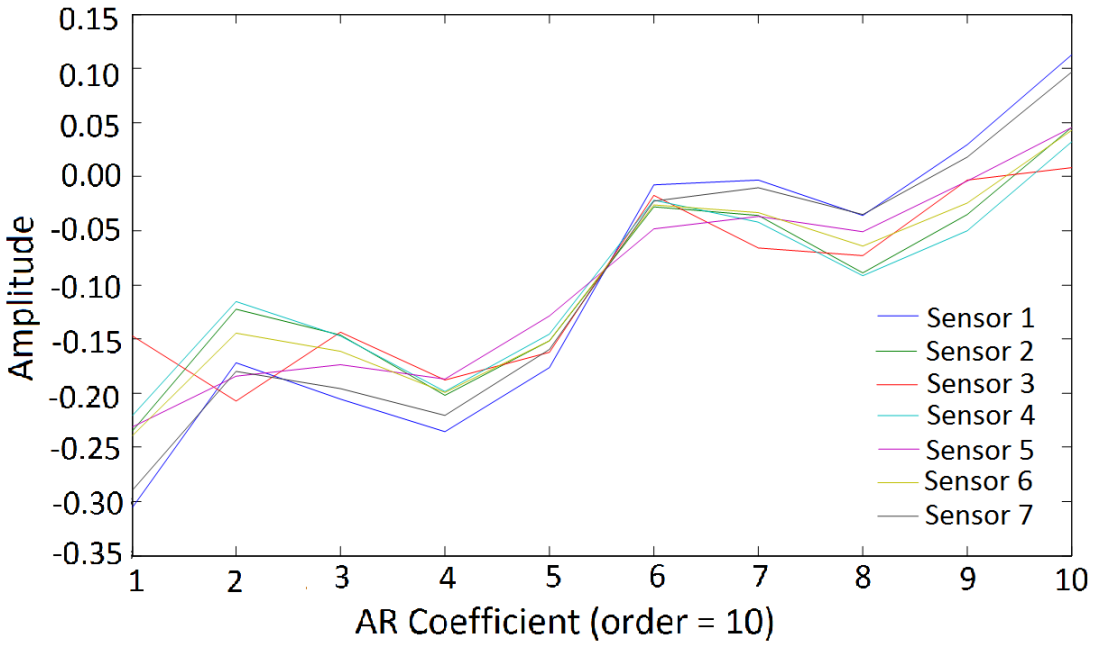


Figure 4. Example of the parameter space representation using AR coefficients (p=10) for a segmented BW signal (0.0013 – 0.2500 s). The computed amplitude (y-axis) of each AR coefficient (x-axis) from the seven sensors for a single impact of pole P1 is shown.

The computationally efficient Burg method estimates AR coefficients from the complex-valued reflection coefficient sequence, based on a least squares criterion, while satisfying the Levinson-Durbin recursion (18). The order of the AR model is always an unknown integer that needs to be estimated from the data. Selection of the AR model order (p) for noisy signals represents a trade-off between increased resolution and decreased prediction error variance of AR coefficients. The Akaike Information Criterion (AIC) method (18) was used to determine the optimal AR model order. The AIC determines the model order by minimising the information theoretic function of the form:

$$AIC[p] = N\ \ln(\tau) + 2p, \qquad (2)$$

where p is the model order, N is the number of data samples given by the stress wave segment size multiplied by the sampling frequency, and τ is the variance estimate of the white noise input to the AR model for order p.

For the timber pole data, the model order is incrementally increased from p=10 to p=60 for the BW excitation method, as shown in Figure 5 (a). In the figure, the average AIC response curve tends to approach an optimum within a range of orders $32 \leq p \leq 36$. For the LW excitation method, the average AIC response curve tends to approach an optimum within a range of orders $9 \leq p \leq 14$, as shown in Figure 5 (b). The optimal range of AR model orders for building feature vectors lies just beyond the 'turning-point or knee-point' of the AIC response curve according to the following principles. Firstly, higher model orders provide a diminishing advantage as their variance estimate values are typically with <1-2% of the knee-values. So using higher AR orders to construct the feature vectors will provide no real benefit in the classification step. Secondly, using a model with larger order will result in significantly higher dimensional feature vectors that will confound machine learning algorithms. Thus, using AR orders located just beyond the knee point (towards the asymptotic minimum of the values shown) provides the simplest description while maintaining the salient features present in the data. In addition, it is likely that the variance estimate values (y-axis) will increase as the model order increases beyond the values shown in the figures.

Parametric methods for signal transformation and the AIC method have been used in similar studies including the detection of structural damage in the presence of operational and environmental variations using vibration-based damage identification procedures (19).

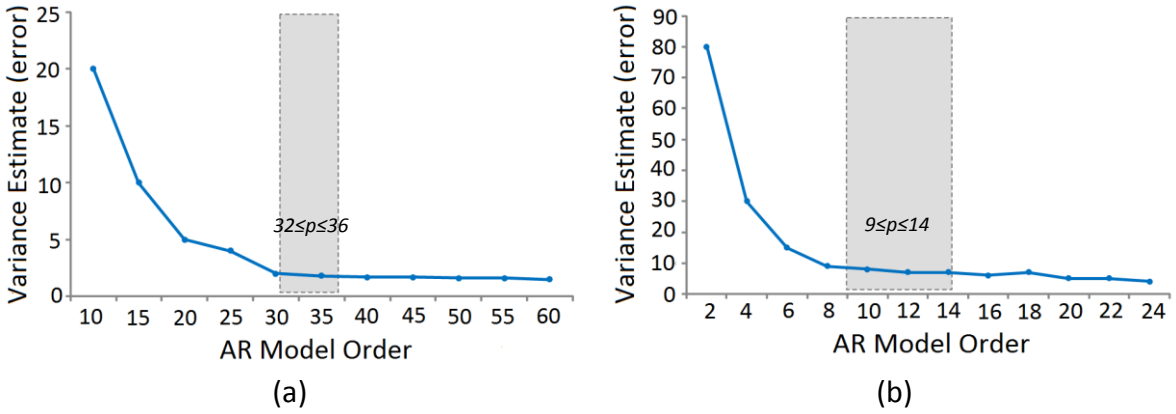


Figure 5. Average AIC model order response for (a) BW and (b) LW excitation methods. The range of model orders to the right of the 'turning- or knee-point', as the graphs tends asymptotically towards the x-axis for the values shown, are the critical orders used to construct feature vectors that summarise the data in the most compact fashion, while retaining the salient features of the data.

AIC: Akaike information criterion; BW: bending wave; LW: longitudinal wave.

Feature vector encoding for classification learning

The production of a single feature vector was the final stage of transforming the GW signal data into a form that is amenable to the input environment of the machine learning algorithms for classification learning. In classification learning, the learning scheme is presented with a set of classified (labelled) examples from which it is expected to learn a way of classifying unseen examples. As such, a static feature vector was encoded and partitioned to produce training and testing vectors for single-step binary classification. A single instance of the

feature vector was encoded for each of the five impact experiments, which were performed on each of the in-situ timber utility poles. Each instance that provides the input to machine learning is characterised by the scalar AR values on a fixed, predefined set of features or attributes. Based on the parametric representation, each instance is a concatenation of the *k*-autoregressive coefficients computed from the segmented time-domain stress wave signal for each of the seven sensors fitted to the test pole as shown in Figure 6. The AR coefficients were computed using the Burg AR method and a single feature vector was created from the resulting coefficients.

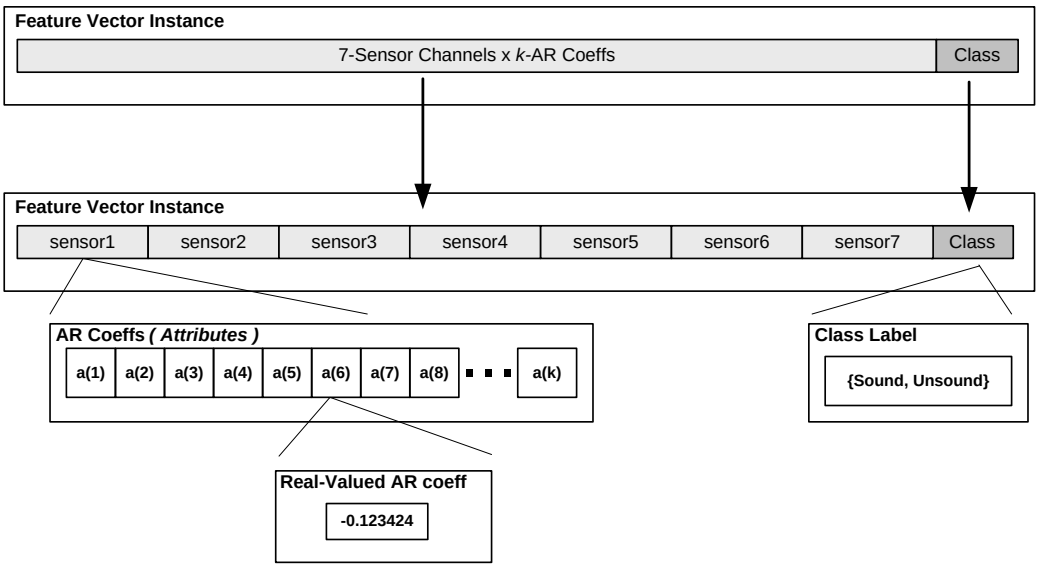


Figure 6. A single feature vector instance (row) based on the AR parametric representation. A single instance typically comprises of k-autoregressive coefficients (attributes) per sensor channel encoded as scalar value and an associated class label.

Training and testing strategy for classification learning

Training and testing sets were derived for Leave-One-Out (LOO) cross-validation supervised classification. LOO cross-validation is simply *n*-fold cross-validation, where *n* is the total number of instances in the feature vector. We produced a training set $\left(X \in \mathbb{R}^{i \times j}\right)$ consisting of (*i*) instances of length (*j*), which were presented to each of the ML algorithm in the training phase. Similarly, the testing set $\left(Z \in \mathbb{R}^{k \times j}\right)$ contained only (*k*) unseen instances, also of length (*j*). As such, the number of instances in the training and testing sets fully covered the instances contained in the feature vector ($n=i+k$).

In this paper, we encoded each instance for a single impact experiment applied to a timber utility pole. Since there are five impact experiments for each pole, there are five instances corresponding to a single pole as illustrated in Figure 7. Hence, the LOO cross-validation method requires that all five instances associated with the same timber utility pole be left out of the training set (X) and used only in the testing set (Z) as unseen data.

All instances associated with the test poles were in turn left out, and the learning scheme was trained on all the remaining instances. It was judged by its correctness of prediction on the

remaining instance - one or zero for success or failure, respectively. The results of all *n* judgments, one for each member of the dataset, were averaged, and that average represents the final error estimate. Based on the averaged *n*-fold prediction (classification) accuracy results, this approach attempts to provide a measure of the generalisation capability of the machine learning algorithm and encoding technique combination. The classification accuracy is defined as:

$$\frac{Number\ of\ correctly\ predicted\ data}{Number\ of\ total\ testing\ data} \times 100\% \qquad (3)$$

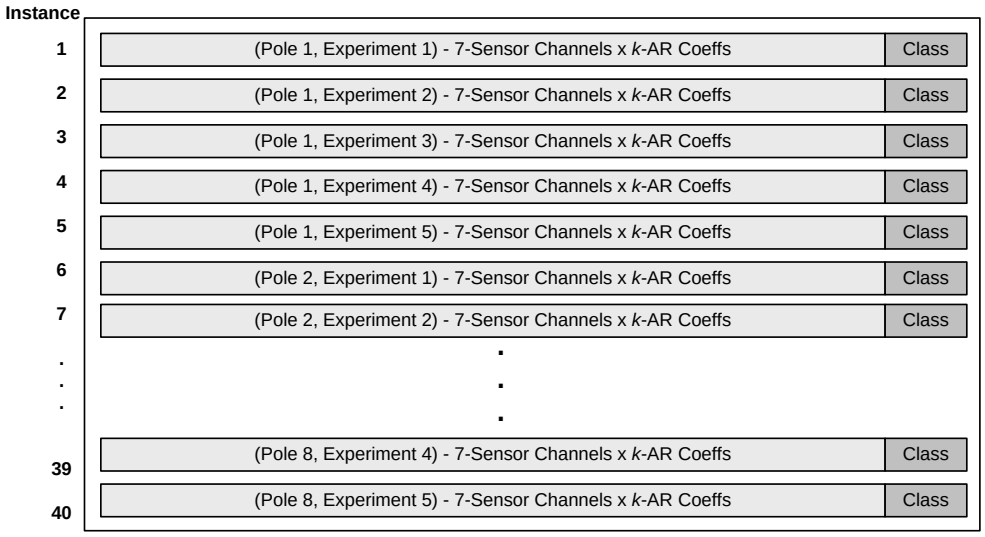


Figure 7. A single feature vector is derived from the concatenated AR coefficients for each sensor/experiment/pole combination. Testing and training sets are derived separately from the final feature vector for LOO cross-validation classification experiments.

However, there is one disadvantage to LOO cross-validation, apart from the computational expense. By its very nature, it cannot be stratified. It actually guarantees a non-stratified sample. Stratification involves getting the correct proportion of instances in each class into the testing set, which is impossible when the test set contains only a single example (20).

Machine learning algorithms

In this paper, we employ four widely used state-of-the-art machine learning algorithms to perform single-step supervised binary classification of the AR-encoded feature vector. Each of the machine learning (ML) algorithms have very different underlying mathematical formulations, but the learning procedure has a common sequence. The goal of ML algorithms is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes.

The first phase for each of the ML algorithms is the supervised training phase. This is performed in a single step, where the training set $\left(\mathrm{X} \in \mathbb{R}^{i \times j}\right)$ is presented to each of the ML algorithm separately. Each of the ML algorithms computes a model from the instances contained in the training set to a different feature space, where the learnt structure is

represented. Once an ML algorithm has been suitably trained on the (i) instances the remaining (k) unseen instances are used as the testing set $\left(Z \in \mathbb{R}^{k \times j}\right)$. The testing set is presented to each of the ML algorithms for the purpose of predicting the binary classification into one of two classes – damaged (1) or undamaged (0). For completeness, four different ML algorithms are briefly described.

Support vector machine (LibSVM). The defining feature of support vector machines (SVM) (21, 22) is the use of linear models to implement nonlinear class boundaries. This is achieved through the transformation of the instance space (AR-encoded feature vector in our case) using a nonlinear mapping into a new space. A linear model computed in the new space can represent a nonlinear class decision boundary in the original space (20). SVM and indeed LibSVM are based on an algorithm that computes a special type of linear model called the maximum-margin hyperplane. For example, a two class dataset is imagined whose classes are linearly separable by a hyperplane in instance space that classifies all training instances correctly. The maximum-margin hyperplane gives the greatest separation (w_{opt}) between the class clusters as shown in Figure 8 (a). Similarly, in Figure 8 (b), the hyperplane H_3 maximises the margin between the two class clusters, while H_2 does not maximise the boundary distance between the class clusters and H_1 does not completely separate the classes at all.

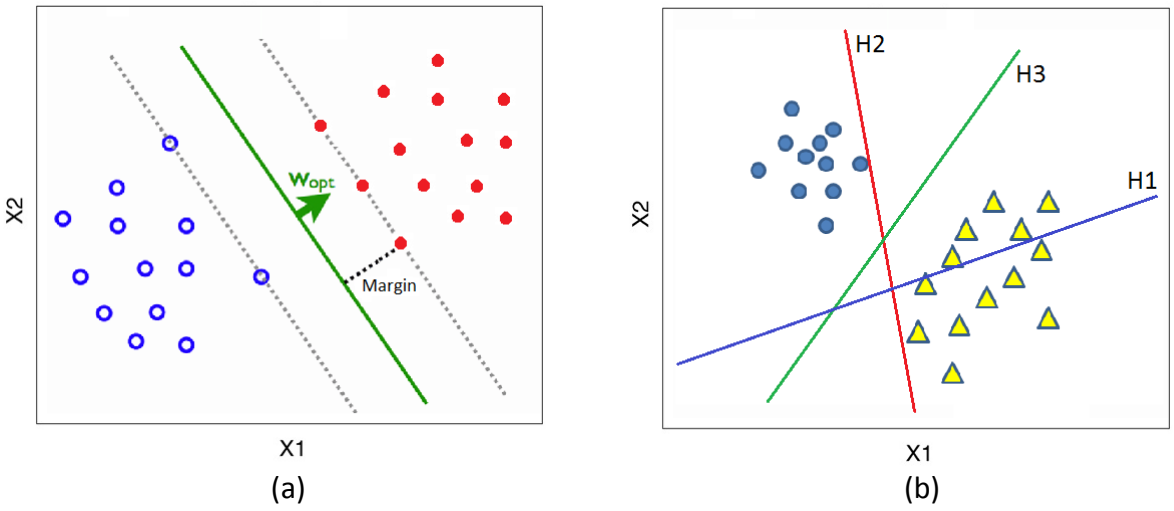


Figure 8. (a) A maximum-margin hyperplane forming a linearly separable decision boundary between two classes (23). (b) Three hyperplanes H1, H2, and H3. H1 does not separate the two classes; H2 separates but with a very small margin between the classes; H3 separates the two classes with a much larger margin than H2 and presents the maximum-margin hyperplane (24).

The instances that are closest to the maximum-margin hyperplane (the ones with minimum distance) are called the support vectors. Each class in instance space must always contain at least one support vector, which lies on the convex hull of a set of points enclosed by a convex polygon. As such, the maximum-margin hyperplane is the perpendicular bisector of the shortest line connecting the class convex hulls.

A hyperplane which can linearly separate two classes can be written as $x = w_0 + w_1 a_1 + w_2 a_2$, where a_1 and a_2 are the attributes and three weights w_i which need to be computed during the training phase. However, this equation can be expressed in general terms of the support vectors themselves:

$$x = b + \sum_{i=1}^{l} \alpha_i y_i a(i) \cdot a \quad (4)$$

where y_i is the class label of the training instance *a(i)*. The learning algorithm computes the parameters *b* and α_i during the training phase (21, 22). Here, *a(i)* are the support vectors and *a* represents the vector of test instances. Finally, *b* and α_i are parameters that determine the hyperplane. Finding the support vectors for the training instances and calculating the parameters (*b* and α_i) is a constrained quadratic optimisation problem resulting in potentially very large and dense data structure called the *Q* matrix.

Above, we have only presented the linear class boundaries for two classes. Now suppose the transformed space is high-dimensional so that the transformed support vectors and test instances have many attributes. According to Equation (4), each time an instance is to be classified the dot product *a(i)·a* with all support vectors must be recalculated, which is computationally expensive in the high-dimensional space produced by the nonlinear mapping. By using a kernel function to compute the dot product before the nonlinear mapping is performed greatly reduces the computational complexity. The high-dimensional version of Equation (4) is simply:

$$x = b + \sum_{i=1}^{l} \alpha_i y_i (a(i) \cdot a)^n \quad (5)$$

where *n* is the number of factors in the transformation (21, 22). Because of the equivalence of Equation (4) and Equation (5), the dot products can be computed in the original low-dimensional space and the problem becomes computationally feasible.

Training an SVM requires the solution of a potentially very large quadratic programming (QP) optimisation problem. This constrained QP problem arises when the SVM algorithm computes the support vectors for the training instances and calculates the parameters (*b* and α_i). The LibSVM implementation uses a decomposition method to iteratively solve the dense matrix arising from the constrained QP problem. The decomposition method modifies only a subset of the dense matrix per iteration. This subset of variables, denoted as the working set, leads to a smaller optimisation sub-problem. LibSVM uses two tricks called shrinking and caching for the decomposition method. To save the training time, the shrinking technique tries to identify and remove some bounded elements during the decomposition iterations, so a smaller optimization problem is solved. In addition, caching is an effective technique for reducing the computational time of the decomposition method. LibSVM can use available memory (called kernel cache) to store some recently used elements for the large *Q* matrix. As a result, some of the kernel elements may not need to be recomputed.

Any function $K(x_i,x_j)$ is a kernel function if it can be expressed as $K(x_i,x_j) = \Phi(x_i) \cdot \Phi(x_j)$, where Φ is a predefined function that maps an instance into a higher (maybe infinite) dimensional space as illustrated in Figure 9.

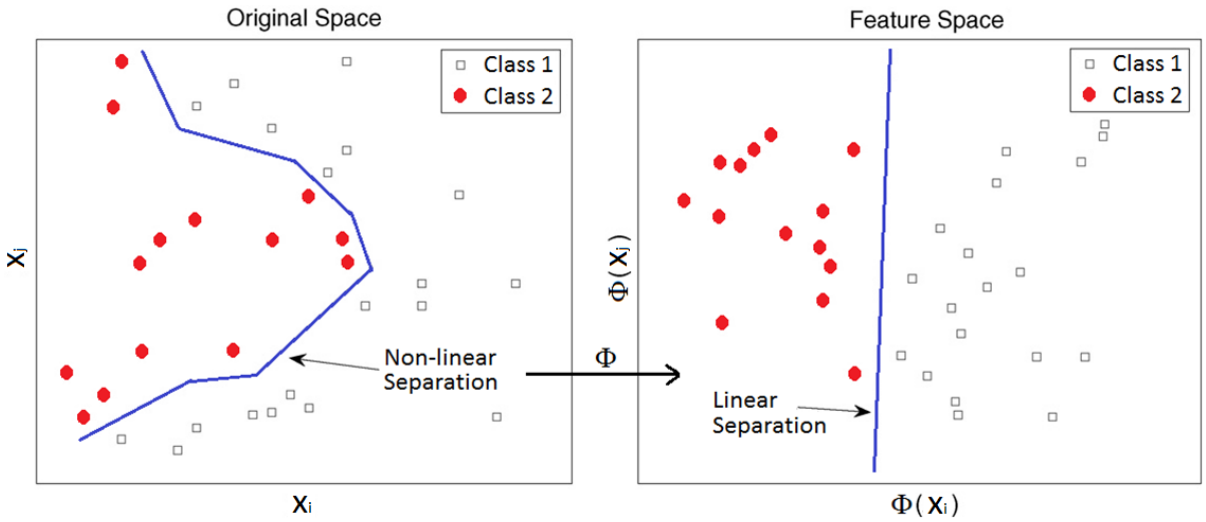


Figure 9. Illustration of kernel function Φ providing a nonlinear decision boundary through its mapping between spaces (adapted from (25)).

Hence, the kernel function represents a dot product in the feature space created by the function defined by Φ. Though new kernels are being proposed by researchers, LibSVM includes four popular kernels:

- linear: $K(x_i, x_j) = x_i^T x_j$.
- polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$.
- radial basis function (RBF): $K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0$.
- sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$.

In this study, we experimented with the linear, polynomial ($3 \leq d \leq 6$), and the radial basis function (RBF) kernels implemented with LibSVM. For a complete description of the LibSVM library for SVM refer to (26).

Sequential minimal optimisation (SMO). The SMO algorithm (27) is a computationally efficient algorithm for training SVM classifiers. As with LibSVM, training a SVM requires the solution of a potentially very large QP optimisation problem. The SMO algorithm partitions the QP problem into much smaller QP problems and solves them analytically at every step, avoiding the time-consuming single-shot numerical QP optimisation, which involves a large matrix computation in the SVM inner loop.

SMO is conceptually simple, easy to implement, is sometimes faster and has better scaling properties than the SVM algorithms which rely on the standard chunking algorithm to optimise the Lagrange multipliers at each step in training. The main advantage provided by SMO comes from its ability to analytically calculate the smallest possible optimisation problem, which consists of two or three Lagrange multipliers at each step. Hence, the entire inner iteration due to numerical QP optimisation is avoided, unlike the other implementations of the SVM algorithm. As a result, SMO is significantly faster for both linear and non-linear kernels.

For a complete description of the SMO algorithm, including experimental results on real-world problems and benchmarking against other SVM-based algorithms refer to (27).

Bayesian Network (BayesNet). A BayesNet is a probabilistic graphical model over a set of variables $\left(Z \in \mathbb{R}^{k \times j}\right)$ that forms a network structure *B*, which is a probabilistic directed acyclic graph (DAG) consisting of nodes and edges. The Bayesian network *B* represents a probability distribution using the DAG. The Bayesnet performs the classification task which consists of classifying a variable $y = x_o$ called the class variable given a set of variables $x = x_1 \ldots x_n$, called attribute variables. The classifier maps an instance of *x* to a value of *y* (h: $x \rightarrow y$). The classifier is learned from a dataset $\left(X \in \mathbb{R}^{i \times j}\right)$ consisting of samples over (x, y). The learning task consists of finding an appropriate Bayesian network structure *B* given the dataset *X* over *Z*.

The Bayesian network structure consists of nodes representing Bayesian random variables and edges representing conditional dependencies. Nodes which are not connected represent conditionally independent variables. Each node is associated with a probability function that takes a set of values for the parent node and provides a probability of the variable represented by the node.

To use a BayesNet as a classifier both inference and learning algorithms are required. We use a simple probability distribution P(Z), which is represented by the Bayesian network structure. This is equivalent to finding the conditional probability distribution $P(y|x)$.

The learning algorithm requires two steps: first learn the structure of the network and then learn the associated probability tables. We use local score metrics to learn the network structure of the BayesNet. The quality measure of the given network structure is a maximisation problem, based on minimum descriptor length, information and other criterion such as AIC and BIC (20). This permits calculation of a score for the global network derived from the local score of individual network nodes. As such, local search methods can be used to help solve the optimisation problem. We use the *K*2 local hill climbing search algorithm.

For a complete description of the Bayesian Network Classifiers and *K*2 search algorithms used for finding solution to the local score metrics optimisation problem refer to (28).

Gaussian Processes (GP). In this study, GP (29) are used for the supervised learning and probabilistic classification, which is the problem of learning input-output mappings (training set (X) -> testing set (Z)) from an empirical data set. Although, the GP are used exclusively for the task of classification in this paper they are equally useful for regression tasks.

To paraphrase Rasmussen, a Gaussian process is a generalisation of the Gaussian probability distribution. Whereas a probability distribution describes random variables which are scalars or vectors, a stochastic process governs the properties of functions. One way of thinking of a Gaussian process is as an infinite-dimensional generalisation of the multivariate normal distribution. As such, one can loosely think of a function as a very long vector, each entry in the vector specifying the function value $f(x)$ at a particular input x. If one asks only for the properties of the function at a finite number of points, then inference in the Gaussian process

will give the same answer if one ignores the infinitely many other points, as if one would have taken them all into account. And these answers are consistent with answers to any other finite queries one may have. One of the main attractions of the Gaussian process framework is precisely that it unites a sophisticated and consistent view with computational tractability and minimal parameter tuning.

In this paper, we employ the polynomial kernel $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$ and radial basis function (RBF) kernels $K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0$. No optimisation of the GP hyper parameters is performed prior to classification and a fixed noise term (s_n = 1.0) is used. For a complete description of Gaussian Processes refer to (29).

EXPERIMENTAL RESULTS AND ANALYSIS

This section presents the experimental results of the LOO training and testing regime (as described in section "Training and testing strategy for classification learning") using the state-of-the-art machine learning algorithms BayesNet, support vector machine, sequential minimal optimisation and Gaussian processes as presented above.

The experimental setting included feature vectors containing the parameter space representation of the stress wave signal data for eight timber poles for the LW excitation and six timber poles for the BW excitation. For the BW data, results are presented for the full data measurement time of 0.5 s and a reduced segment time of 0.25 s. Results for the supervised learning classification experiments based on the BW excitation method are presented in Figure 10 to Figure 13, while results for the experiments based on the LW excitation method are illustrated in Figure 14 and Figure 15.

In multi-class classification experiments, the result of a single test set is typically analysed using the two-dimensional confusion matrix with a row and column allocated to each class or category. Each matrix element shows the number of test examples for which the observed class is the row and the predicted class is the column. Good results correspond to large numbers running down the main diagonal and small, ideally zero, off-diagonal elements. The results based on classification accuracies do not provide information about the distribution of predicted class, rather a convenient result of overall classification accuracy. Hence, we employ the Cohen Kappa Statistic to measure the agreement between predicted and observed classes of a dataset, while correcting for an agreement that occurs by chance. It is a method to describe the distribution of predicted classes in the confusion matrix. The Kappa value ranges from [0, 1] with a value of 1 indicating perfect prediction with all values lying on the central diagonal (20).

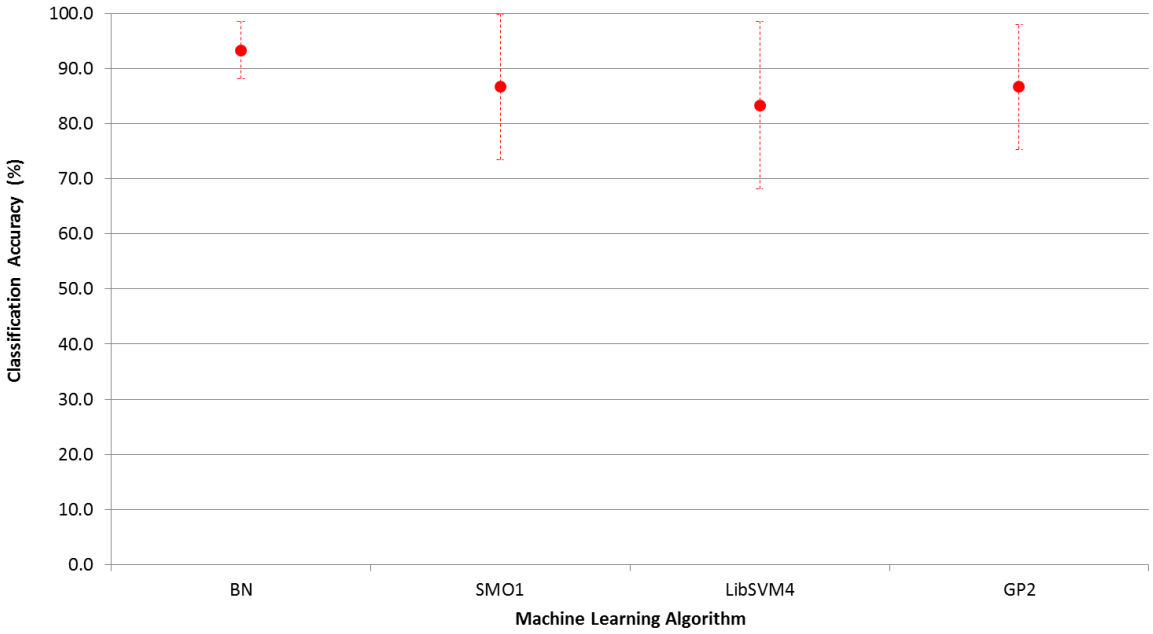


Figure 10. Mean ± SD classification accuracy of the parametric-based feature vector (p=50) for BW data (segment time = 0.5 s).

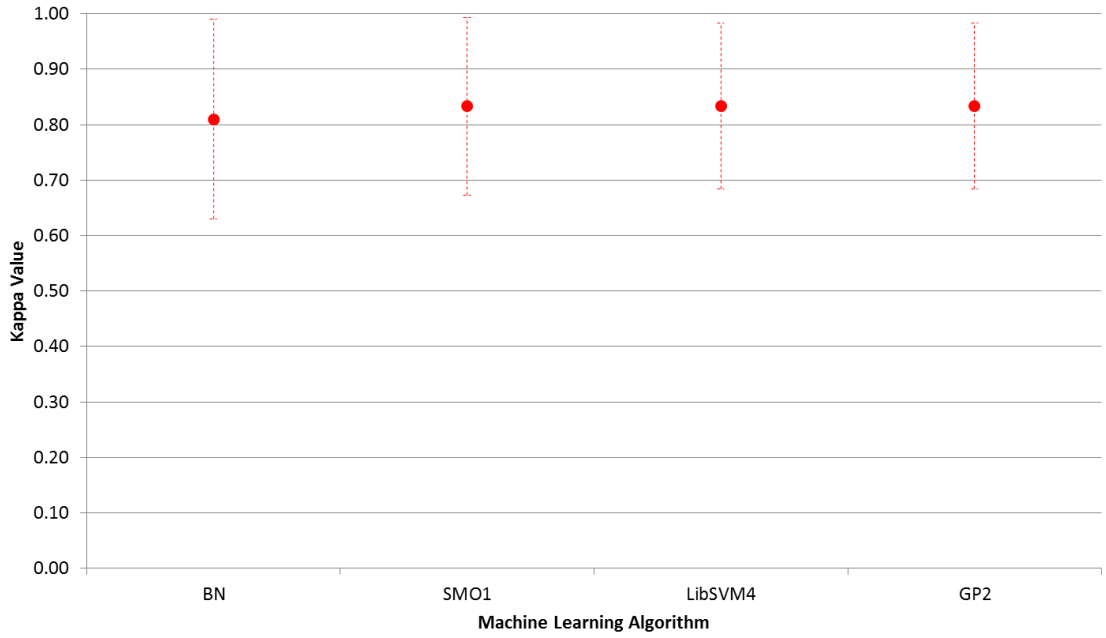


Figure 11. Mean ± SD Cohen Kappa Statistic of the parametric-based feature vector (p=50) for BW data (segment time = 0.5 s).

According to a paired Student's *t*-test (α=0.05, n=8), for the mean classification accuracy, the BayesNet (BN) classifier does not provide a statistically significantly better classification accuracy than the support vector machine (SMO1), sequential minimal optimisation (LibSVM4) or Gaussian processes (GP2) classifier (p-values = 0.33, 0.13, 0.20 respectively)

using the feature vector (p=50) with BW excitation data and a segment time of 0.5 s (see Figure 10 and Figure 11.

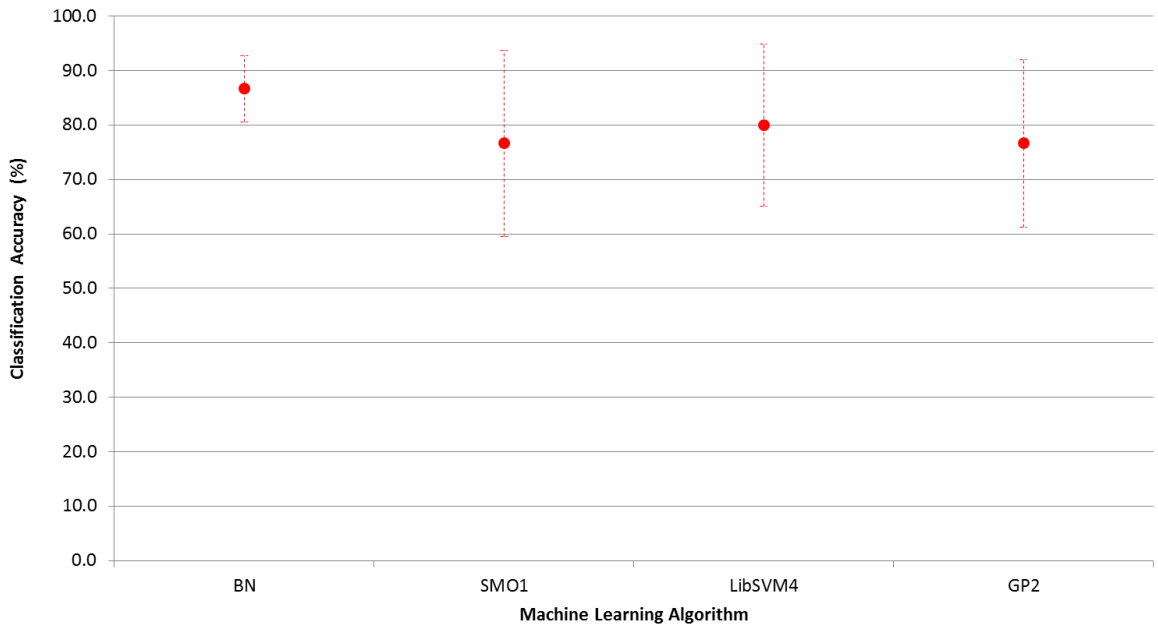


Figure 12. Mean ± SD classification accuracy of the parametric-based feature vector (p=36) for BW data (segment time = 0.25 s).

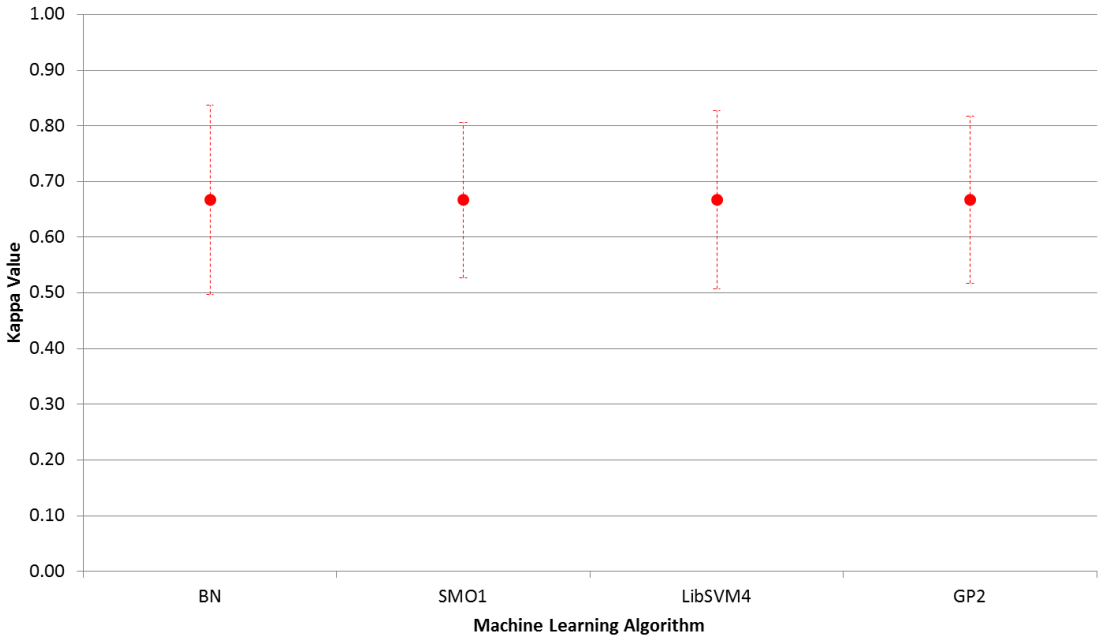


Figure 13. Mean ± SD Cohen Kappa Statistic of the parametric-based feature vector (p=36) for BW data (segment time = 0.25 s).

For a segment time of 0.25 s (BW excitation), the BN classifier does also not provide a statistically significantly better classification accuracy than the SMO1, LibSVM4 or GP2

classifiers (p-values = 0.17, 0.29, 0.14 respectively) using the feature vector (p=36), as depicted in Figure 12 and Figure 13. In addition, no statistically significantly higher classification accuracy and associated Kappa statistic are achieved for any machine learning algorithm employed in this study, for a window size of 0.5 s and AR order of (p=50) for BW excitation over a window size of 0.25 s and AR order of (p=36).

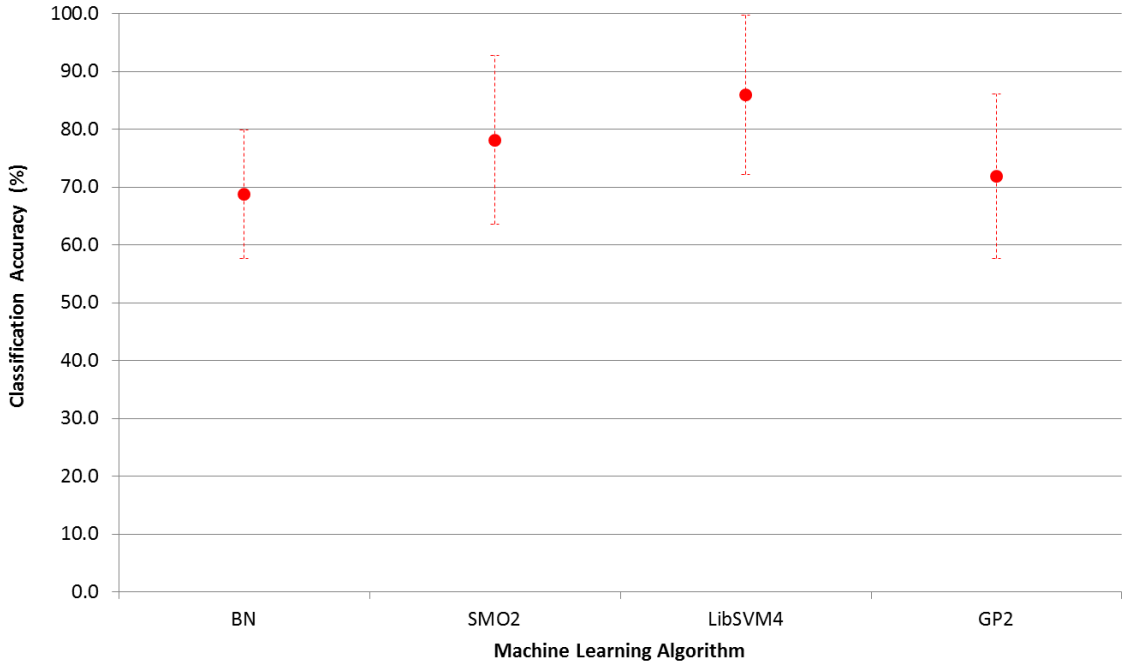


Figure 14. Mean ± SD classification accuracy of the parametric-based feature vector (p=10) for LW data (segment time = 0.25 s).

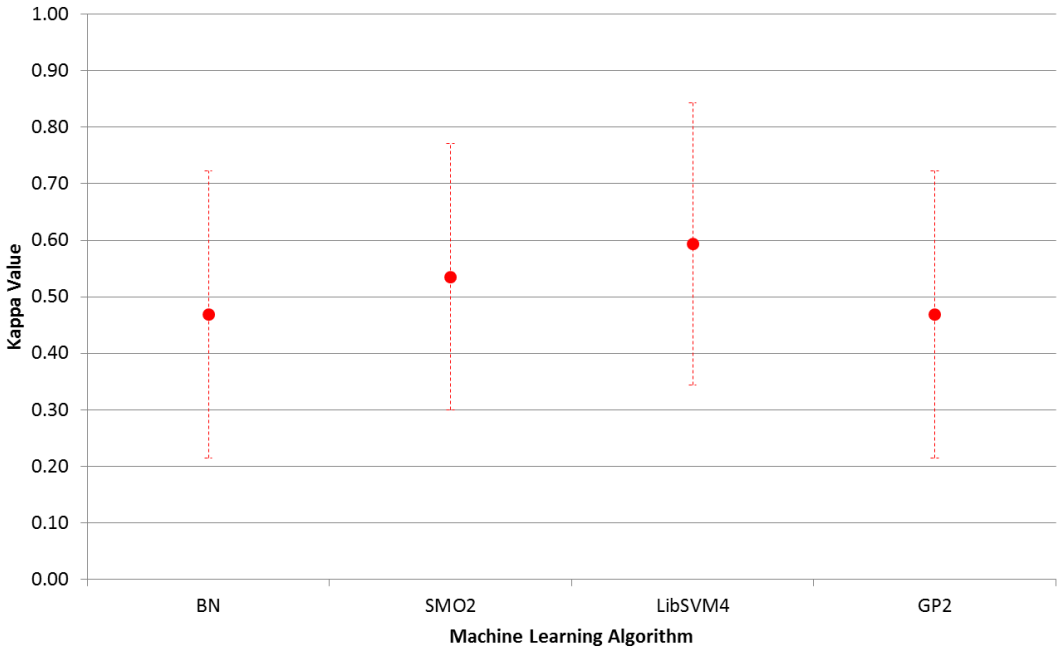


Figure 15. Mean ± SD Cohen Kappa Statistic of the parametric-based feature vector (p=10) for LW data (segment time = 0.25 s).

For data from LW excitation (segment time = 0.25 s), the LibSVM4 classifier does provide a statistically significantly better result classification accuracy than the BN and GP2 classifiers (p-values = 0.02, 0.05 respectively) using the feature vector (p=10), according to a paired Student's t-test (α=0.05, n=8). Furthermore, there was no statistically significantly better classification accuracy and associated Kappa statistic for a window size of 0.5 s and AR order of (p=10) for LW excitation (results not shown here).

SUMMARY AND CONCLUSION

This paper presented a statistical signal processing approach based on parametric methods coupled with a supervised machine learning techniques to perform classification results for the structural health monitoring (SHM) of in-situ timber utility poles based on guided wave (GW) propagation. The proposed method utilises an innovative multi-sensor testing system that captures wave response signals along a sensor array and it applies machine learning algorithms for the pattern recognition and classification of statistically transformed measurement signals to evaluate the soundness of a pole including its embedded section. Using leave-one-out (LOO) cross validation, it was found that using an autoregressive (AR) model order within the range of $46 \leq p \leq 50$ and segment time of 0.5 s did provide a classification accuracy of 93.3±6.0% with 'excellent agreement' between the observed and predicted classes as indicated by the kappa statistic of 0.81±0.18, for the bending wave (BW) excitation method. For the longitudinal wave (LW) excitation method, it was found that an AR model order within the range of $9 \leq p \leq 14$ and segment time of 0.25 s did provide a classification accuracy of 85.7±10.8% with 'substantial agreement' between the observed and predicted classes as indicated by the kappa statistic of 0.61±0.25.

It is not the aim of this paper to suggest the 'best' machine learning algorithm to use in this domain. However, the results calculated from a two-sample Student's *t*-test with 95% confidence and sample size (n=8) for the mean classification accuracies indicate there is no statistically significant difference between the machine learning algorithms used in this study. The only exception was the result for the LW excitation, where the SVM-based algorithm did provide a slightly better result.

ACKNOWLEDGEMENTS

The authors wish to thank the Centre for Built Infrastructure Research (CBIR), Faculty of Engineering and Information Technology, University of Technology Sydney (UTS) for supporting this project. PowerCore is thanked for enabling the testing of the timber utility poles and Dr Fook Choon Choi and Amir Zad are thanked for conducting the field experiments.

Funding

This project was supported by the Australian Research Council and our industry partner Ausgrid, through Linkage Project [LP110200162]. Additional support was provided through the Industry and Innovation Project scheme of the Faculty of Engineering and Information Technology, University of Technology Sydney (UTS).

REFERENCES

1. Marsh E. *Composites in infrastructure - Building new markets*. 1st ed. Amsterdam: Elsevier Science Ltd, 2000.
2. Kent H. Options for the continued supply of power poles. Summary of research for the Power Poles and Cross Arms Committee, Energy Networks Association of Australia. , In: *Presentation at the Australian Wood Pole Resources Workshop*, 2006, Brisbane, Australia: Queensland Department of Primary Industries and Fisheries, Energy Networks of Australia.
3. Tanasoiu V, Miclea C and Tanasoiu C. Non-destructive testing techniques and piezoelectric ultrasonics transducers for wood and built in wooden structures. *J Optoelectron Adv M* 2002; 4(4).
4. Nguyen M, Foliente G and Wang X. State of the practice and challenges in non-destructive evaluation of utility poles in service - Advances in non-destructive evaluation. *Key Eng Mater* 2004; 270-273.
5. Stewart AH. Wood poles life span: What you can expect. IEEE/PES Expo Edition. 1996; 20.
6. Paquet J. Etude Vibratoire des Pieux en Beton: Reponse Harmonique et Impulsionelle, Application au Controle. In: *Annales de L'Institut Technique du Batiment et Des Travaux Publics*. 1968; 21(245): 789-803.
7. Steinbach J and Vey E. Caisson Evaluation by Stress Wave Propagation Method. J Geotech Eng-ASCE. 1975; 101(GT4): 361-78.
8. Van Koten H and Middendorp P. Interpretation of results from integrity tests and dynamic load tests, In: *Proceedings of the International Seminar on the Application of Stress Wave Theory in Piles*, Stockholm, Sweden, 1980, pp. 217-232.

9. Davis AG and Dunn CS. From theory to field experience with the non-destructive vibration testing of piles. In: *Proceedings of Institution of Civil Engineers*; 1974.
10. Ross RJ, Brashaw BK, Wang X, White RH and Pellerin RF. Wood and timber condition assessment manual. Madison, WI: Forest Products Society; 2004.
11. Qian Y and Mita A. Damage identification of structures based on pattern classification using limited number of sensors. In: *Proceedings of the 5th International Workshop on Structural Health Monitoring*, 2005; Stanford, CA, USA.
12. Olson LD, Jalinoos F and Aouad MF. Determination of unknown subsurface bridge foundations. Washington, DC, USA: Transportation Research Board, National Research Council, 1995.
13. Kasal B, Lear G and Tannert T. In situ assessment of structural timber. 1st ed. Kasal B, Tannert T, editors. Netherlands: Springer, 2010.
14. American Concrete Institute. Nondestructive test methods for evaluation of concrete in structures. Farmington Hills, Michigan, US: Report ACI 228.2R-98, 1998.
15. Kim YR, Ranjithan SR, Donato PJ and Murphy CM. Nondestructive evaluation of the structural condition of timber piles. Final report, North Carolina Department of Transportation, FHWA/NC/2000-004; 2000. 574.
16. Chen S and Kim YR. Condition assessment of installed timber piles by dispersive wave propagation. *J TRB*. 1996; 1546: 112-20.
17. Qian Y. Nondestructive evaluation of structural condition of timber piles using stress wave techniques. Raleigh, NC, USA: North Carolina State University, 1997.
18. Marple SL. Digital spectral analysis with applications. 1st ed. Oppenheim AV, editor. Englewood Cliffs, N.J: Prentice-Hall; 1987.
19. Figueiredo E, Park G, Farrar CR, Woden K and Figueiras J. Machine learning algorithms for damage detection under operational and environmental variability. Struct Health Monit 2011; 10(6).
20. Witten IH, Frank E and Hall MA. Data mining: Practical machine learning tools and techniques. 3rd ed. Gray J, editor. San Francisco: Morgan Kaufmann; 2011.
21. Cortes C and Vapnik VN. Support-vector networks. Machine learning. Kluwer Academic Publishers-Plenum Publishers. 1995; 20(3): 24.
22. Boser BE, Guyon IM and Vapnik VN. A training algorithm for optimal margin classifiers. In: *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*. 1992; 144-152.
23. Huang TSD, Rajaram S, Chang EY, Mandel MI, Poliner Graham E and Ellis DPW. Active learning for interactive multimedia retrieval. *Proceedings of the IEEE*. 2008; 96(4): 648-67.
24. Antonio Carlos Gay T. SVM classifiers – Concepts and applications to character recognition. In: Ding X (ed) *Advances in Character Recognition*. 1st ed. InTech, 2012.
25. Cerqueira AS, Ferreira DD, Ribeiro MV and Duque CA. Power quality events recognition using a SVM-based method. *Electr Pow Syst Res* 2008; 78(9): 1546-1552.
26. Chang C-C and Lin C-J. LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol* 2011; 2(3): 27.
27. Platt JC. Advances in Kernel methods - Support vector learning. Schölkopf B, Burges C, A. Smola, editors: MIT Press; 1999. 185-208.
28. Cooper G and Herskovits E. A Bayesian method for the induction of probabilistic networks from data. *Mach Learn* 1992; 9(4): 309-347.
29. Rasmussen CE and Williams C Gaussian processes for machine learning (Adaptive computation and machine learning): MIT Press; 2005.