Machine Learning Analysis of Features Extracted from Time-Frequency Domain of Ultrasonic Testing Results for Wood Material Assessment

Mohsen Mousavi[a], Amir H Gandomi *[a], Damien Holloway[b], Adam Berry[a], Fang Chen[a]

[a]*Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW 2007, Australia.*
[b]*School of Engineering, University of Tasmania, Hobart 7005, Tasmania, Australia*

Abstract

In this paper, a systematic machine learning strategy is proposed to classify wood properties based on a contact ultrasonic testing results. As such, several aspects of the wood material including the type of wood (hardwood or softwood), the direction of an ultrasonic test with respect to the growth rings of the wood, and, whether the wood is damaged or intact are investigated. As a pre-processor, the Variational Mode Decomposition technique is applied to the nonlinearly modulated ultrasonic signals, and the center frequencies of the decomposition results are taken as features for Machine learning Algorithms (MLAs). Then with each of the MLAs, hyperparameter settings were optimised and technical aspects of the feature engineering are discussed. Best results were achieved using Ensemble classifiers, SVM, and KNN using three features/decomposition. To explore the physics behind the nonlinear problem, the relative false discovery rate obtained from the confusion matrix associated with applying the MLAs is proposed as a metric. We show that different features are capable of exploring different aspects of the problem better. The techniques of this paper can be applied to quality assessment of wood materials. The paper also demonstrates the capability of MLAs in exploring some physics of such problems. Moreover, billets with natural imperfections harvested from the site Collie in WA, Australia are classified to demonstrate the applicability of the proposed approach in real world problems. The result of 94% 5-fold cross-validation accuracy indicates the effectiveness of the proposed approach.

Keywords: `Ultrasonics, Variational Mode Decomposition, Machine learning, Nondestructive damage testing, Wood`

Email address: `Corresponding authors: gandomi@uts.edu.au` (Amir H Gandomi *)

1. Introduction

Different techniques have been exploited for non-destructive damage testing of wood materials, the most common of which are ultrasonics [1], radiography [2] and thermography techniques [3]. Ultrasonic testing has been the most widely used approach for assessment of wood material properties [4, 5, 6, 7, 8, 9], mainly because ultrasonic tests are both non-intrusive and inexpensive [10], and ultrasonic waves are highly sensitive to wood mechanical properties [11].

There are many factors that can affect the wood quality adversely, either during production and/or while in-service. It is known that if standing trees in plantations are not pruned in time, they will undergo a process of natural self-pruning which, in turn, will leave knots in the trunk [12]. These knots are regarded as defects in the wood, making it less useful for industrial application. The two main factors affecting the quality of in-service wood materials post construction are biological and physical processes [13]. The former occurs due to decay, fungi, and termites, while the latter results from climatic condition variations such as rain and sun cycles.

Damage assessment in wood sections is therefore of great importance in different industries. Such assessment may be required at different stages in the product life-cycle. For instance, mechanised harvesting prompts wood damage monitoring in order to evaluate wood properties in standing trees [14]. Damage detection in in-service wooden poles is of great importance to ensure the functionality of these structures [15, 16, 17]. Moreover, preservation of wooden architectural heritage has received a great deal of attention during the past decades [18, 19, 20, 21].

There are two types of ultrasonic devices that are commonly used for non-destructive damage assessment of wood sections, namely contact and non-contact ultrasonic devices. The use of non-contact ultrasonics for NDT has been reported widely in the literature. This include electromagnet ultrasonics (EU) [22], laser ultrasonics (LU) [23], and air-coupled ultrasonics (ACU) [24, 25, 26, 27]. However, there are some limitations in terms of applying a non-contact ultrasonic device for assessment of wood properties. For instance, EU devices do not perform well in non-conductive mediums. On the other hand, LU devices are relatively expensive, where industrial laser ultrasonic systems can range from 0.5 to 2 million U.S. dollars [28]. It is also known that while ACU devices are effective when used for assessment of low density materials such as wood [26], the quality of ACU signals is usually poor, which will further demand more advanced signal processing techniques to be used to process signals obtained from ACU tests [25]. However, the effectiveness of these signal processing approaches are yet to be borne out by further experiments. Therefore, CU seems to be the most promising technique for quality assessment of wood materials due to its relatively cheap price and good transmissibility to the wood materials.

However, there are still some challenges with using contact ultrasonic devices, which makes further study of the application of this technique for assessment of wood materials crucial. Some of these challenges stem from the necessity of using a couplant gel to fill the gap between the transducer/receiver and the surface of the wood section, to ensure proper transmissibility of the ultrasonic waves into the

wood material. For instance, the pressure applied to the transducer/receiver can squeeze some of the gel out of the gap, which will further affect the quality of the test, especially if the pressure is not maintained and the gel does not completely bridge the gap. Moreover, any misalignment or vibration of the transmitter or receiver can also affect the quality of the test. Therefore, the results of the test may be highly affected by these factors.

There are generally two approaches that can be taken when it comes to studying wood using ultrasonic techniques. The first one, that involves the study of more complex underlying physics, is the guided wave technology (GWT). The guided wave ultrasonic testing (GWUT) allows the inspection of an elongated member through guiding ultrasound by the member's boundaries. The proper probing frequency in this technique is dependent on the size and thickness of the studied member, ranging from 20 to 100 kHz [29]. Moreover, the method mainly relies on advanced mathematical modeling, graphically presented in plots, called dispersion curves. However, due to the complexity of guided waves, the solutions of the wave equation for structures composed of orthotropic materials, such as cylindrical wooden poles, are not yet at hand [30]. As such, GWUT techniques are not fully capable of interpreting wave patterns extracted from field measurements conducted on timber structures. This is mainly due to the fact that the GW propagation in wood structures, such as timber poles, is relatively incomprehensible [31].

Conventional ultrasonic techniques, on the other hand, monitor the time of flight and/or velocity of ultrasonic bulk waves propagated in materials for evaluation of their properties [32, 33, 34, 35]. However, since Wood is an orthotropic material [36], many of its properties vary with direction [37, 38], resulting in independent mechanical properties in three mutually perpendicular axes, namely 1) the longitudinal direction, parallel to the fiber/grain, 2) the radial direction, normal to the growth rings of the wood, and 3) tangential direction, tangent to the growth rings [39]. On the other hand, due to the change of the microfibril angle from one type of wood to another, this diversity of mechanical properties is also a function of wood species [40]. Therefore, there are many uncertainties involved with using conventional techniques for assessment of wood material properties. For instance, even though the wood species can usually be determined by visual inspection, it is very difficult to identify after testing, just from the ultrasonic signal, the direction through which the ultrasonic test was conducted. This will make it hard to distinguish which direction the evaluated mechanical properties correspond to (e.g. tropical woods, [41]).

MLAs provide strong tools for tackling the aforementioned problems associated with using CU devices. As such, the aim of this paper is to use MLAs to solve a supervised learning problem of ultrasonic signal classification derived from an ultrasonic test conducted on wood specimens using a CU device. To that end, we propose to solve a classification problem that classifies wood specimens based on: 1) the type of the wood, 2) state of damage of the wood, and 3) direction of the test with respect to the growth rings of the wood.

To demonstrate these features, tests conducted on softwood and and hardwood, and in the tangential

and radial directions (with respect to the growth rings of the wood), are studied. In addition, both healthy and damaged conditions for each type of wood specimens are considered, with the damage implemented in different directions. In many studies, an artificial defect is synthesised in wood sections to study their effects on wood properties [5, 42, 43, 13], and a similar technique is used in the present paper.

In order to minimise uncertainties resulting from the quantity of applied gel and other test conditions, 50 replicate of each test are conducted. The outcomes of the ultrasonic tests are ultrasonic signals that are nonlinearly frequency-modulated due to their interaction with wood grains and damage geometry [44, 45]. In order to derive some features for the classification problem, Variational Mode Decomposition (VMD), a nonlinear signal decomposition algorithm, is proposed to be employed through decomposing the modulated ultrasonic signals into their constructive oscillatory modes [46]. The center frequencies of the extracted modes are taken as features to be fed into the classification problem.

Data classification problems can be seen as a process of four general steps: 1) data collection, 2) feature extraction, 3) feature classification using machine learning models, and 4) model optimisation with the aim of optimising the hyperprameters of the selected model (Figure 1). However, while often used as 'black boxes', MLAs are believed to be able to provide further insightful information about the physics of the problem beyond solving the classification task. One of the informative output of MLAs is the confusion matrix.

As mentioned above, the main novelty of the present work relies on employing and testing the capability of the variational mode decomposition algorithm for decomposing ultrasonic bulk waves into some intrinsic mode functions, termed IMFs, through which deriving meaningful features are made possible. One key factor of the proposed method is that it uses features derived from the time-frequency domain of the ultrasonic signals. As such, the derived features are independent from the length through which the ultrasonic wave travels. This makes the proposed method successful in health assessment of specimens regardless of their geometry.

Generally, signal decomposition has been proven effective in extracting information from ultrasounic waves. For instance, Najjar et. al [30] employed the EEMDAN algorithm (complete ensemble empirical mode decomposition with adaptive noise) to identify mode reflections and compare two ultrasonic signals obtained from guided ultrasonic tests conducted on wooden poles. The authors concluded that most information was embedded in the first four modes of the decomposition results. The higher modes (IMFs 5 to 10), that correspond to lower frequency ranges compared to the original signal, contained less information. In another study conducted by Mousavi et al. [46], the authors concluded that three decompositions of an ultrasonic bulk wave obtained from a contact ultrasonic test, using VMD, can extract enough information about damage. The higher modes of the VMD algorithm are of higher frequency content and may either present duplicated information or do not carry any information about damage at all.

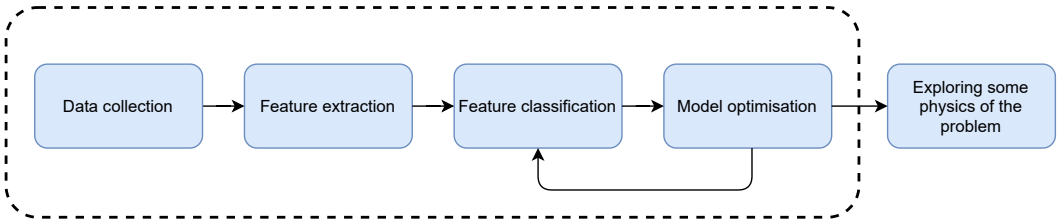


Figure 1: Augmented data classification steps.

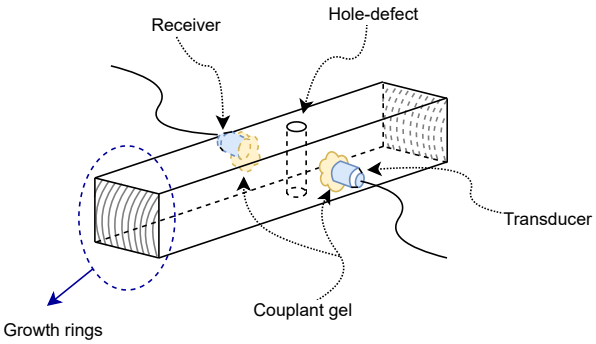
Machine learning algorithms have been used for solving different classification and regression problems of wood inspection using ultrasonic testing. For instance, Nasir et. al [47], employed machine learning algorithms to work out the modulus of elasticity (MOE), and modulus of rupture (MOR) in a UV-degraded wood sections using decision trees. In this paper, a further step is taken in which the confusion matrix of the optimised machine learning results is used to explore some aspects of the physics of the problem. In order to determine which feature addresses which property of the tested specimens better, a metric is proposed that uses the confusion matrix obtained from solving the classification problem using only one of the obtained features. We will show that useful information can be derived from the proposed metric, the details of which are fully discussed in the sections to follow. Further, experimental study of the billets harvested from standing trees with and without natural imperfections demonstrate the effectiveness of the proposed method for classification of wooden specimens based on their health condition.

2. Experimental set-up and measurements

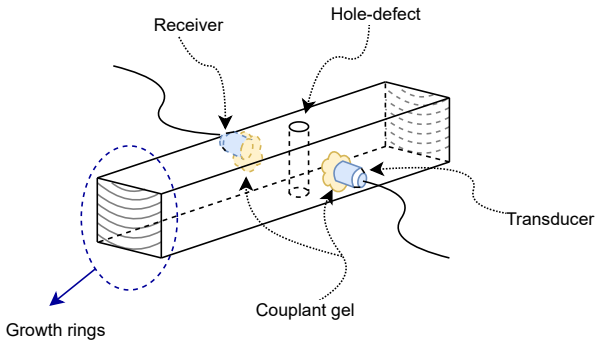
Two types of woods are studied in this paper, a softwood (Radiata Pine, or *pinus radiata*) and a hardwood (Merbau, or *instia palembanica*). There were in total six specimens of each type, of size $300 \times 90 \times 90$ mm^3. Three different testing scenarios were conducted: (1) intact specimens, (2) specimens with a small defect in the form of a hole drilled through them, and (3) specimens with a large defect. In scenario (1), all six intact specimens of each type were first tested in both their radial and tangential directions. In scenario (2), a 6 mm hole, roughly 7% of the cross section of the specimens, was drilled into three out of six specimens in the tangential direction to simulate a small hole-defect. The specimens were then tested in the direction perpendicular direction to the line of defect, i.e. along the radial direction as depicted in Figure 2b. The remaining three specimens of the same type were drilled with a hole of the same size through their radial direction and were tested through the tangential direction. These are illustrated in Figure 2c. In scenario (3), the hole-defects in all defective specimens from scenario (2) were enlarged to 13 mm diameter, roughly 14% of the cross section of specimens, and were tested in a similar manner as in scenario (2).

The meteorological conditions of the room upon testing were as follows: there were limited temperature fluctuations–the temperature ranged from 20–22°C. The average humidity level was around 60%. The ultrasonic testing device was a Pundit PL200 (Figure 2a), in which a 54 kHz transducer was used

(a) Ultrasonic device (Pundit PL 200)



(b) Radial test



(c) Tangential test

Figure 2: Ultrasonic test experimental set-up.

to transmit a sinc-like probing P-wave (compression wave). The pulse repetition frequency (PRF) was 5 Hz. The receiver recorded the transmitted and modulated ultrasonic waves (due to its interaction with wood irregularities) at a sampling frequency of 10 MHz. Similar sampling frequencies were applied in other works such as [47].[1] Couplant gel (Proceq Ultraschall-Koppelpaste) was used to fill the gap between transducer/receiver and the specimen surfaces in order to ensure proper transmissibility of the ultrasonic waves into and out of the wood materials. However, there were uncertainties regarding the amount of gel used and the level of the pressure applied to the transducer/receiver by hand. As such, 50 replicates of each ultrasonic test were conducted for each configuration. Table 1 shows the number of tests conducted on specimens of different type and health condition, and in different directions.

As can be seen from Table 1, there were 12 different configurations of specimens, encompassing different wood spcies (Merbau "M" and Pine "P"), health condition (intact "I", small damage "S",

[1]The readers are referred to [48] for further details.

Table 1: The number of tests per sample regarding 12 different classes.

Specimen and the test type	**# class**	**# tests**
M_r^I	1	300
M_r^S	2	150
M_r^L	3	150
M_t^I	4	300
M_t^S	5	150
M_t^L	6	150
P_r^I	7	300
P_r^S	8	150
P_r^L	9	150
P_t^I	10	300
P_t^S	11	150
P_t^L	12	150

and large damage "L"), and the direction of the test (radial "r" and tangential "t"). Considering the replication of samples (6 for each intact condition and 3 for each damaged condition) and the repetitions of the test (50 for each configuration), there were, in total, 2400 test results available.

Next we explain how some damage sensitive features (DSF) can be derived from each measurement using the advanced signal decomposition technique, Variational Mode Decomposition (VMD).

3. Feature extraction using Variational Mode Decomposition

The transmitted ultrasonic signals are hypothesised to be nonlinearly modulated due to their interaction with wood grain and damage. One way to extract information out of these signals is, therefore, to decompose them into their constructive modes. To that end, however, a nonlinear decomposition algorithm should be used. Therefore, we propose to use the Variational Mode Decomposition (VMD) algorithm to perform the decomposition, and to take center frequencies of all modes as features. The details follow.

VMD is a time–frequency signal decomposition algorithm that solves a variational optimisation problem to decompose a given complex signal $S(t)$ into a specified number K of oscillatory modes, $\{u_k\} = \{u_1, u_2, ..., u_K\}$, denoted Intrinsic Mode Functions (IMFs) [49]. VMD can be also used to denoise the signal while decomposing it into its noise-free modes. In the case that denoising is not intended, the sum of all IMFs perfectly reconstructs the original signal.

Each IMF can be both frequency and amplitude modulated, hence is of the form

$$u_k(t) = A_k(t)\cos(\phi_k(t))\,, \tag{1}$$

where $A_k(t)$ and $\phi_k(t)$ are the time varying amplitude and phase of the k^{th} IMF. The instantaneous frequency $\omega_k(t) = d\phi_k(t)/dt$, is also time dependent, but is constrained by the extraction algorithm to vary significantly more slowly than $\phi_k(t)$.

Despite the time dependent property of the phase, each IMF is narrow-band, i.e. the frequency variation occurs within a limited frequency range. Therefore, IMFs extracted from VMD are considered as monocomponent. As such, the center frequency of each IMF can well characterise that IMF and, therefore, is taken as a feature in this paper. There can be K center frequencies (features) extracted from K IMFs as $\{\omega_k\}$, where $k \in \{1, ..., K\}$.

The constraint variational optimisation problem of VMD is,

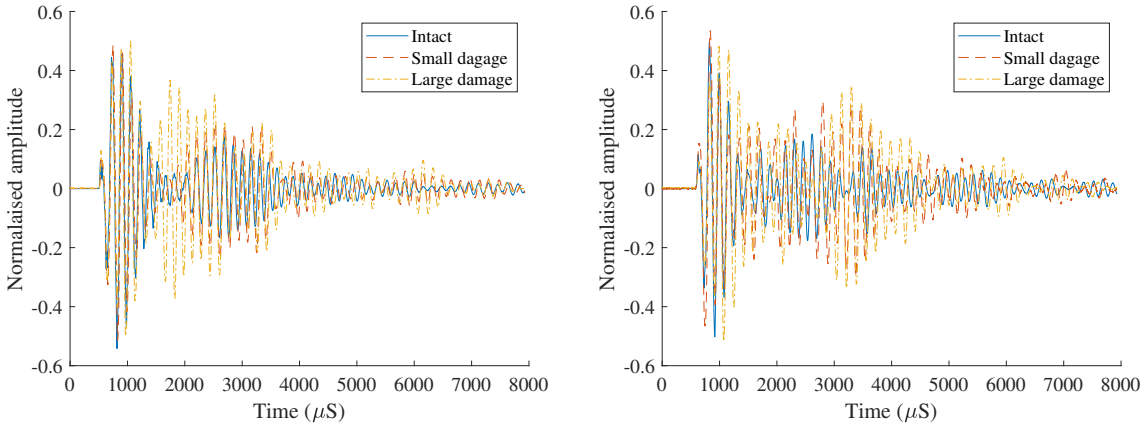
$$\min_{\{u_k\}\ \&\ \{\omega_k\}} \sum_k \left\| \partial_t \left(\delta(t) + \frac{j}{\pi t} * u_k(t) \right) e^{-j\omega_k t} \right\|^2 \quad \text{s.t.} \quad S(t) = \sum_k u_k(t), \tag{2}$$

where $*$ is the convolution operator, j is the imaginary unit, and δ is the Dirac distribution. The solution to the minimization problem of (2) is the saddle point of its augmented Lagrangian in a sequence of iterative sub-optimizations, i.e. alternate direction method of multipliers (ADMM) [49],

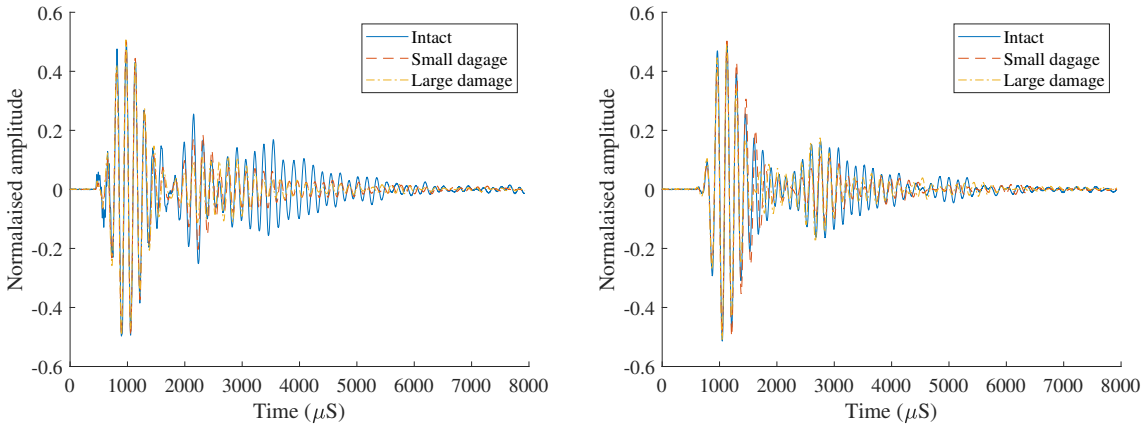
$$\begin{aligned} \mathcal{L}\left(\{u_k\}, \{\omega_k\}, \lambda\right) := \alpha \sum_k \left\| \partial_t \left(\delta(t) + \frac{j}{\pi t} * u_k(t) \right) e^{-j\omega_k t} \right\|_2^2 \\ + \left\| S(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), S(t) - \sum_k u_k(t) \right\rangle. \end{aligned} \tag{3}$$

VMD is thus a parametric decomposition algorithm and, therefore some parameters need to be specified prior to running the decomposition task as follows:

1. K: is an integer that determines the number of IMFs into which the signal is to be decomposed. K IMFs also means K center frequencies or features. Therefore, by varying K, different numbers of features can be extracted. In this paper, several K values have been trialled.
2. α: is the quadratic penalty term and is a denoising factor, a larger value of which results in less noise in the IMFs.
3. τ: is the time step, which determines how quickly the Lagrangian multiplier accumulates the reconstruction error. One may set τ to a small number such as 0.1 (as suggested by proposers of VMD [50]) should an exact reconstruction be intended. In such a case, the effect of α will be neglected. Otherwise, one can set τ to zero, where the reconstruction is not strictly enforced, but encouraged in a least-squares sense. In this paper, a low-pass filter is used for denoising, therefore, to make the effect of α irrelevant, τ is set to 0.1.
4. ϵ: is the error tolerance, and controls the convergence of the algorithm. A value of $\epsilon = 10^{-5}$ is chosen in this paper.
5. *init*: which initialises ω either as either zero ($init = 0$), uniform ($init = 1$), or random ($init = 2$). Based on the findings in [51], the method of initialising center frequencies has little effect on the decomposition results. Therefore, $init = 0$ in this paper.



(a) M_r^I (b) M_t^I



(c) P_r^I (d) P_t^I

Figure 3: Examples of normalised ultrasonic signals obtained from different specimens with various health condition.

6. *DC*: is a Boolean variable and determines whether or not the first IMF is set and kept at DC (a mode with zero center frequency). Since the center frequency of each IMF (including the first IMF) is taken as a feature in this paper, it is important that *DC* be set to 0 (false), otherwise one can simply ignore the first IMF.

First, each ultrasonic signal $S(t)$ is normalised through shifting by its mean value (μ), and scaled by the difference between its maximum and minimum values as follows [48]:

$$\overline{S}(t) = \frac{S(t) - \mu}{max - min} \tag{4}$$

Figure 3 shows examples of normalised recorded ultrasonic waves obtained from testing different types of specimens with various health condition. It can be seen from the plots that the normalised amplitude of the ultrasonic signals recorded off the Merbau specimens expands as damage progresses in the specimens. However, this trend is completely reversed as for Pine specimens, where the normalised amplitude of the recorded ultrasonic waves shrinks as the specimens are a more damaged.

The normalised signals are then low-pass filtered with a cutoff frequency of 300 kHz. The signals thus obtained are decomposed into K IMFs and the center frequencies of all IMFs are taken as features.

Figure 4 shows examples of ultrasonic signals measured from ultrasonic tests conducted on the four different intact configurations, namely Merbau and Pine, tested respectively in their radial and tangential directions.

4. Importance of each feature

There is always a question about the effectiveness of the extracted features in the classification problems. Therefore, we perform a principal component analysis (PCA) on each different number of features/decompositions (different K values) to explore the importance of each feature and the necessity of using a higher number of decomposition. It is known that the most important variables are those that are correlated with the lower order principal components, especially PC_1. Accordingly, features that are not correlated with any PC or are correlated with the lower dimensions are less important and can be simply omitted from the dataset. Note that the dataset (matrix of extracted features) needs to be standardised prior to running PCA[2].

The principal components of the standardised feature matrix $\mathbf{\Omega}_{m\times k}$ of rank $L \leq \min\{m, k\}$, whose rows are the observations, and columns are the obtained features (center frequencies) for each observation, are obtained from the singular value decomposition (SVD),

$$\mathbf{\Omega} = \mathbf{P}\Delta\mathbf{Q}^{\mathrm{T}}\,, \tag{5}$$

where $\mathbf{P}$ is an $m \times L$ matrix of left singular vectors, $\mathbf{Q}$ is a $k \times L$ matrix ($\mathbf{Q}^{-1} = \mathbf{Q}^{\mathrm{T}}$) of right singular vectors, and Δ is the diagonal matrix of singular values. The principal components of $\mathbf{\Omega}$ are then the columns of the matrix of factor scores, $\mathbf{F}$, representing the distance of the projected observations on the principal axes to the origin, and obtained as

$$\mathbf{F} = \mathbf{P}\Delta\,. \tag{6}$$

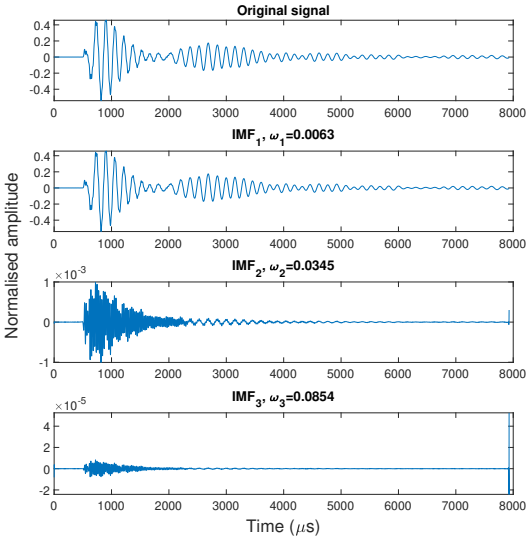
Since multiplying $\mathbf{\Omega}$ by $\mathbf{Q}$ gives the projected values of the observations on the principal components, $\mathbf{Q}$ can be also interpreted as a projection matrix. This can be shown as

$$\mathbf{F} = \mathbf{P}\Delta = \mathbf{P}\Delta\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{\Omega}\,\mathbf{Q}\,. \tag{7}$$

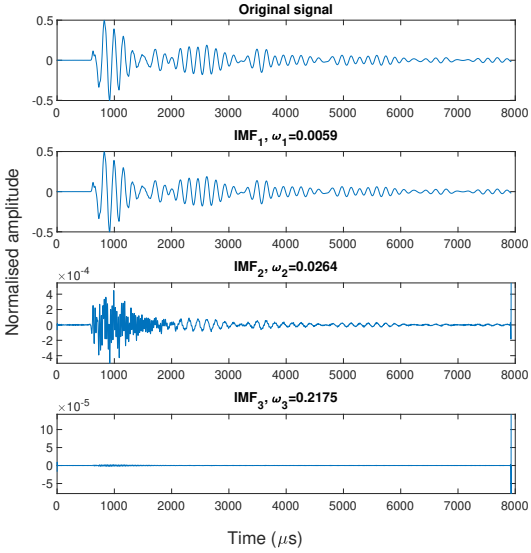
The importance of an observation with regards to a component can be represented by the ratio of the squared factor score corresponding to the observation, divided by the sum of the squared factor scores associated with the component. This ratio is called the "contribution" of the observation to the component [52]. In other words, the contribution of the observation i to the component l, denoted as $\mathrm{ctr}_{i,l}$, is obtained as

$$\mathrm{ctr}_{i,l} = \frac{f_{i,l}^2}{\sum_i f_{i,l}^2}\,. \tag{8}$$

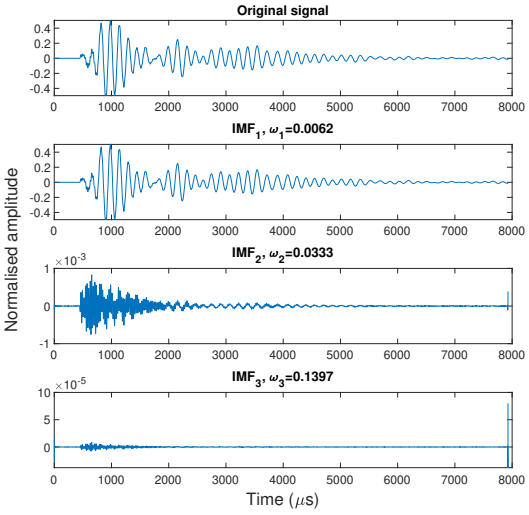
[2]Each column of the feature matrix has to be shifted by its mean value and scaled by its standard deviation to obtain a standardised feature matrix.



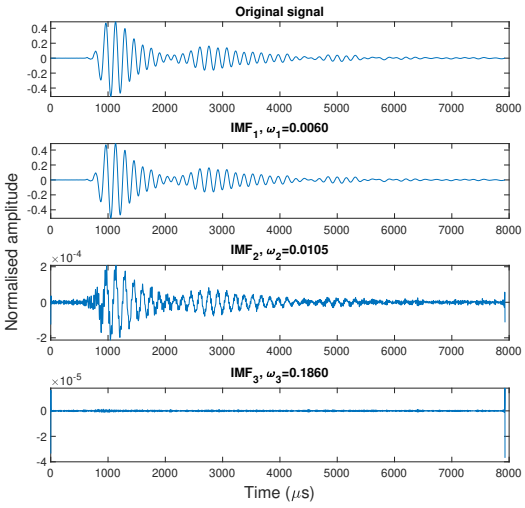
(a) M_r^I



(b) M_t^I



(c) P_r^I



(d) P_t^I

Figure 4: Examples of decomposed signals for different types of intact wood and test direction. The spike appearing at the end of the time interval in some IMFs is due to the end effects. Note that ω above each signal is expressed in MHz, so for example the transmitter probing wave frequency, 54 kHz, would correspond to $\omega = 0.054$. Note also that the analysed samples are considerably shorter than pulse repetition period of 2×10^5 μs.

(a) 3 features/decomposition modes

(b) 4 features/decomposition modes

(c) 5 features/decomposition modes

(d) 6 features/decomposition modes

Figure 5: Contribution percentage of each feature to the principal dimensions corresponding to different principal components when the number of decomposition K varies.

Similarly, the importance of a component to a given observation, termed squared cosine, is obtained as follows:

$$\cos^2_{i,l} = \frac{f^2_{i,l}}{\sum_l f^2_{i,l}}\,. \tag{9}$$

In the PCA framework, the information shared between a component and a variable is estimated through their correlation value which is called "loading". This can be obtained through calculation of the squared values of the entries of matrix of loadings $\mathbf{Q}$ obtained from the singular value decomposition of the standardised observation matrix $\mathbf{X}$. For further information, the readers are referred to [52].

Figure 5 shows the contribution of each feature to the principal components as a percentage, for values of $k = 3$, 4, 5 and 6 representing the number of features or decomposition modes. It can be seen that in each case all the features (center frequency ω_i) contribute almost equally to the dimension corresponding to PC_1. Although the percentage of the contributions of the features to PC_2 is not as uniform, we can still confidently conclude that all features contribute significantly to the variability of the

feature space, and therefore, a higher number of decomposition modes does not reduce the importance of higher order features. Therefore, using a different number of features/decomposition modes might improve the performance of the classification task. As such, several values of $k \in \{2, \ldots, 12\}$ are trialled in different machine learning models in the next section to solve the classification problem.

5. Solving the classification problem

The MATLAB machine learning toolbox is used to solve the classification problem, exploiting six different types of MATLAB supervised MLAs: decision trees, support vector machines (SVM), naive Bayes classifiers, nearest neighbor classifiers, discriminant analysis, and ensemble classifiers. Reiterating, different numbers of features/decomposition modes, $k \in \{2, \ldots, 12\}$, are used as inputs. 5-fold cross-validation is run on the dataset using each of the above supervised MLAs and the obtained 5-fold cross validation accuracy results are presented in the graphs of Figure 6.[3]

As can be seen from the figure, the classification results can either improve or worsen when the number of features/decomposition modes is increased. Specifically, the classification results generally improve with more features when using support vector machines, discriminant analysis, and ensemble of classifiers. The results of the decision trees seems also to improve slightly as the number of features increases. On the other hand, the opposite results are achieved for Gaussian naive Bayes and coarse KNN. Note that we demonstrated in Section 4 that increasing the number of decomposition modes does not compromise the importance of features. Therefore, we conclude that the optimum number of features is only correlated with the type of MLA used for solving the classification problem.

Each MLA has its own hyperparameters that need to be optimised, in addition to the number of features, which itself is a hyperparameter.[4] There are two types of hyperparameters: 1) hyperparameters that require selection of a category, such that between two or more types of functions, and 2) hyperparameters that require specification of a real or integer value selected in an interval between a lower and upper bound. Table 2 lists the tunable hyperparameters associated with each machine learning model, along with the relevant bounds or function type choices. It can be seen that SVM and ensemble classifiers are the most tunable models followed by KNN. The least tunable model is discriminant analysis with only one tunable hyperparameter.

The hyperparameters relating to the best results (based on the number of features used) obtained from each MLA are optimised using a Bayesian optimization algorithm, and the results are presented in Figure 7. The optimised hyperparameters are derived for each case at each iteration. The iteration

[3] We also ran 2-fold cross validation and achieved quite similar results as presented in this paper. However, we presented the 5-fold cross validation results as it is the typical way of dividing datasets into training and test sets.

[4] Note that the number of features/decomposition modes is not incorporated in the hyperparameters tuning problem. Alternatively, we have taken the least number of features/decomposition modes that brings about the best result in each case as the optimum number of decomposition modes. However, one may optimise the number of features/decomposition modes exploited for the classification problem through solving the Bayesian optimisation problem.

corresponding to the set of hyperparameter values that minimises an upper confidence interval of the classification error objective model is taken as the best-point hyperparameter set. The upper confidence interval is defined as the value γ in

$$P(\mu_Q(\mathcal{O}(p)) > \gamma) = \alpha\,, \tag{10}$$

where $\mu_Q(\mathcal{O}(p))$ is the mean value of the Bayesian objective function for the set of parameters p with respect to the posterior distribution Q. The value α indicates the significance level. It is noted that the best-point does not always correspond to the set of parameters that minimises the classification error.

The following sections describe the outcomes for each MLA.

5.1. Decision trees

Figure 6a shows the results obtained when applying the decision trees algorithm. The best and worst results are seen to be achieved by the fine and coarse tree models respectively. The optimum number of features/decomposition modes is 4, for which 99.3% accuracy is achieved using fine trees architecture. In order to further optimise the model hyperparameters, the optimisation algorithm is now run on this case.

The optimisation results of Figure 7a suggest that the best-point of 99.5% accuracy is obtained by selecting the *Twoing rule* as the split criterion, with 198 splits. The number of splits was selected from the range $\{1, \ldots, 2399\}$. It is noted that a different optimum number of splits was obtained each time the optimisation algorithm was run, but the Twoing rule was always suggested as the best split criterion among three possible criteria, namely *Gini's diversity index*, *Twoing criterion*, and *Maximum deviance reduction*.

The finding that the Twoing rule was always the best split criterion is compatible with the fact that the Twoing rule is regarded as more appropriate for data with a large number of different classes – there are 12 in this case. It is known that the Twoing criterion performs equally as well as Gini's diversity index when the target attribute is binary [53]. For multi-class problems the Twoing criterion prefers attributes with evenly divided splits [54].

5.2. Support vector machines

Figure 6b suggests that the SVM model performs well when either the cubic or fine Gaussian SVM is used. Although the medium Gaussian SVM performs relatively well, the worst results are obtained using simpler SVM algorithms, namely coarse Gaussian and linear SVM. Three features/decomposition modes were then used in an optimisable SVM model to optimise the corresponding hyperparameters, and the accuracy achieved for the optimised model was 99.7%.

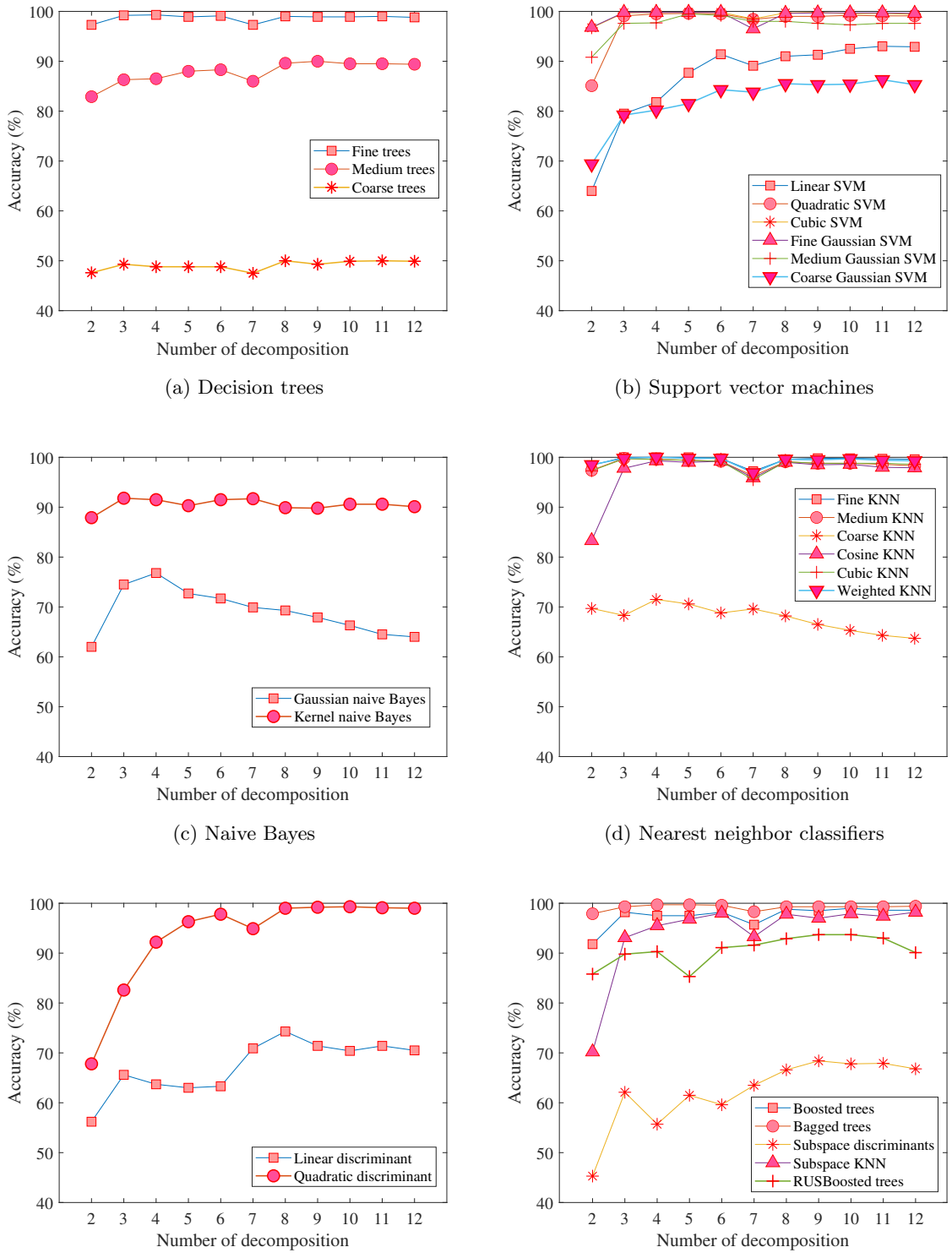
It is noted above that SVM is one of the most tunable MLAs. Tuning of the several required hyperparameters is described following.

Support Vector Machines employ two different strategies to achieve multi-class classification, namely *One-vs-All* and *One-vs-One* approaches. SVMs were initially developed for binary classification, so do

Table 2: Hyperparameters corresponding to each machine learning model in MATLAB. n and n_p are respectively the number of observation and predictors and [.] returns the integer part of a number.

Model	**Optimisable Hyperparameters**	**Type**	**Lower**	**Upper**
Decision trees	- Maximum number of splits	-	1	$\max\{2, n–1\}$
	- Split criterion	- Gini's diversity index, Twoing rule, and Maximum deviance reduction.	-	-
SVM	- Kernel function	- Gaussian, Linear, Quadratic, and Cubic.	-	-
	- Box constraint level	-	0.001	1000
	- Kernel scale	-	0.001	1000
	- Multiclass method	- One-vs-One and One-vs-All.		
	- Standardise data	- True and false.	-	-
Naive Bayes	- Distribution names	- Gaussian and Kernel.	-	-
	- Kernel type (if applicable)	- Gaussian, Box, Epanechnikov, and Triangle.	-	-
KNN	- Number of neighbors	-	1	$\max\{2, [n/2]\}$
	- Distance metric	- Euclidean, City block, Chebyshev, Minkowski (cubic), Mahalanobis, Cosine, Correlation, Spearman, Hamming, Jaccard	-	-
	- Distance weight	-Equal, Inverse, and Squared inverse.	-	-
	- Standardise	- true and false.	-	-
Discriminant analysis	- Discriminant type	- Linear, Quadratic, Diagonal Linear, and Diagonal Quadratic.	-	-
Ensemble classifiers	- Ensemble method	- AdaBoost, RUSBoost, LogitBoost, GentleBoost, and Bag.	-	-
	- Maximum number of splits	-	1	$\max\{2, n–1\}$
	- Number of learners	-	10	500
	- Learning rate	-	0.001	1
	- Number of predictors to sample	-	1	$\max\{2, n_p\}$

not natively support multi-class classification tasks. They instead split the multi-class classification task into multiple binary classification problems. The *One-vs-All* strategy splits the classification task into as



(a) Decision trees

(b) Support vector machines

(c) Naive Bayes

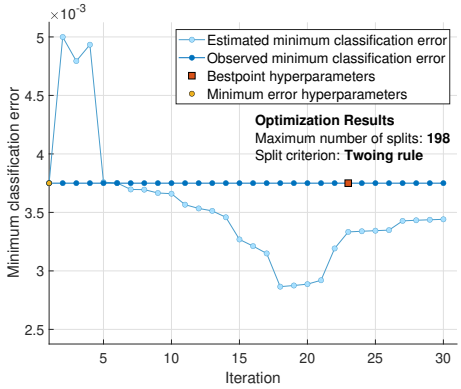
(d) Nearest neighbor classifiers

(e) Discriminant analysis

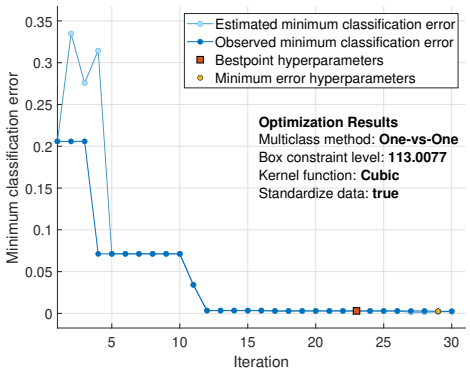
(f) Ensemble classifiers

Figure 6: Accuracy results obtained by using different MLAs with different numbers of features/decomposition modes.

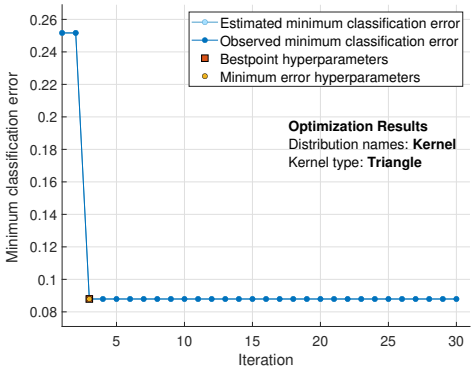
many binary classification problems as there are classes, where, in each problem, one class is reclassified versus the rest of the classes. The alternative, known as a *One-vs-One* classifier, is to distinguish each class from only one other class. If c is the number of classes, there are $\binom{c}{2}$, i.e. $c(c-1)/2$, possible pairs of classes. Each point is then classified according to a majority vote amongst the discriminant functions [55]. The latter method was found to be the optimum classification strategy every time the



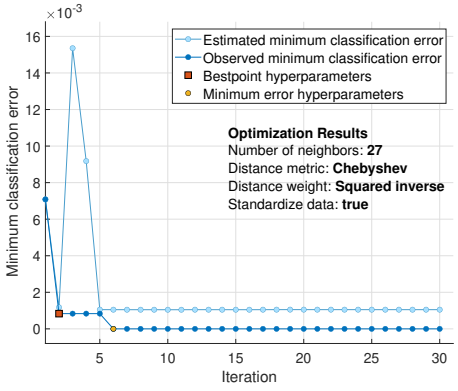
(a) Decision trees



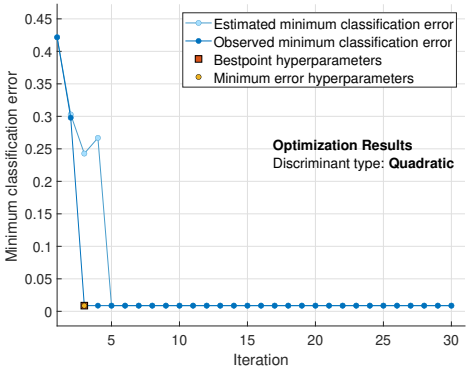
(b) Support vector machines



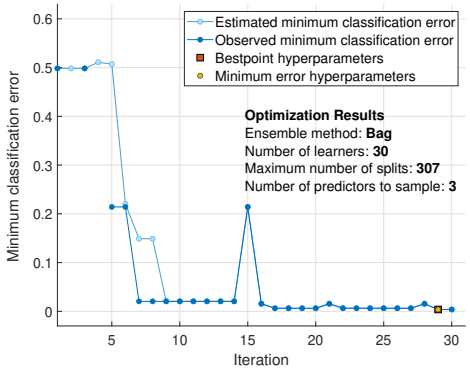
(c) Naive Bayes



(d) Nearest neighbor classifiers



(e) Discriminant analysis



(f) Ensemble classifiers

Figure 7: Optimised machine learning algorithms.

optimisation task is run in the problem of this paper (Figure 7b).

The box constraint level is a parameter that controls the maximum penalty imposed on margin-violating observations. Therefore, it aids in preventing overfitting and has a regularization effect. Its value is selected from the interval $[0.001, 1000]$. Although the optimum value of 113 is shown in Figure 7b, this optimum varied slightly each time optimisation task was rerun.

The default kernel scale, i.e. 1, is kept unchanged since all the features (center frequencies) are in the range 0–0.3, due to the fact that the cutoff frequency is 300 kHz and the sampling interval is 1 μs or 10^{-6} s. Therefore, the inner product of features does not produce any dominant effect coming from one

particular feature. The cubic kernel function is selected as the optimum kernel and data normalisation is selected as *true*.

5.3. Naive Bayes classifiers

The naive Bayes classifier is one of the simplest MLAs, and favors the simplified assumption of independent features. The results of Figure 6c show that it gives less accurate results than almost all the other MLAs. It is also evident from the results that kernel naive Bayes performs much better than the Gaussian naive Bayes. In fact, the Gaussian naive Bayes performs increasingly less accurately as the number of features/decomposition modes increases beyond 4. One reason for this could be due to higher level features (center frequencies of higher order IMFs) not mimicking a Gaussian distribution function, in contrast to the lower order features (center frequencies of lower order IMFs). A kernel distribution fits the distribution of features more appropriately in those cases. Figure 7c shows the most optimum kernel to be the triangle kernel. The optimum number of features/decomposition modes was found to be 3 in this case and a maximum accuracy of 91.2% is obtained through optimising the hyperparameters.

5.4. K-nearest neighbor classifiers

As indicated in Table 2, KNN is one of the most tunable MLAs. KNN finds the q closest points (neighbors) based on their distances to the query point, then votes for the most frequent label associated with those points. In weighted KNN, the weight decreases as the distance increases. The most important hyperparameters for KNN are the number of neighbours q, the metric used for measuring the distances between each query point and its neighbors, and (in weighted KNN) the distance weight method.

Similarly to the SVM, the data normalisation Boolean was selected as *true*. As expected by comparison to the SVM, and confirmed by the results of Figure 6d, the coarse KNN is not an appropriate model for solving the problem of this paper. Here 'coarse KNN' means approaching the upper bound of $n/2$ neighbours, or half the total number of observations (see Table 1), where in this study $n = 2400$. Similarly, 'fine KNN' means approaching the lower bound of 2 neighbours.

In Figure 7d, $K = 27$ was used, being found to be the optimum number of neighbors, and the Chebyshev metric with the squared inverse distance weight was distinguished as the best metric among all the other possible options listed in Table 2. The optimum number of features/decomposition modes in this case is found to be 3 and an accuracy of 99.9% was achieved through optimising the hyperparameters.

5.5. Discriminant analysis

Discriminant analysis is a classification and dimensionality reduction MLA that aims to draw boundaries among classes based on their distributions. These boundaries can be either linear or nonlinear. Therefore, as can be seen from the Table 2, there is only one tunable hyperparameter to optimise. It was found that the *Quadratic* discriminant type gave an accuracy of 99.2% using 10 features/decomposition modes.

5.6. Ensemble classifiers

The last MLA model for solving the problem of this paper is ensemble of classifiers, which uses multiple learners called base learners. Ensemble methods are based on the combination of a set of hypotheses to be used for training, as opposed to using a single hypothesis, as is the case for every base learner individually [56]. As a result, the generalization ability of an ensemble of classifiers is usually much stronger than that of each base learner. Moreover, ensemble learning is able to boost the performance of weak learners[5], to become strong learners that can make very accurate predictions. However, using base learners that are not-so-weak often results in better performance. The resulting trained model is also less prone to overfitting on the training set.

There are different ensemble methods, as listed in Table 2, from which the *Bag* method using 3 features/decomposition modes is found to optimise the current model (Figure 7f). 30 learners were used, a maximum of 307 splits was chosen in the decision tree, and the number of optimum predictors per sample was 3. Accordingly, 99.6% accuracy was achieved by optimising the model. It was noted that all the optimised hyperparameters can vary each time when repeating the optimisation process.

5.7. Summary of classifier comparisons

We have shown that different numbers of features/decomposition modes provide optimum results when applying different MLAs to solve the classification problem of this paper. The results show that SVM, KNN and Ensemble methods provide high accuracy using only three features/decomposition modes, and therefore are chosen as the most effective MLAs for the problem of this paper. However, in order to further analyse the performance of these algorithms in solving the problem of this paper, we propose to add noise to the three derived features/decomposition modes and to solve the classification task with the new noisy features. This will simulate a greater variability in the observations that might exist in an *in situ* environment that is less controlled than the testing regime.

6. Classification using noisy features

In this section, in order to ensure that observations present sufficient variability, we propose to contaminate the extracted features with white Gaussian noise through the following equation,

$$\hat{\mathbf{\Omega}} = \mathbf{\Omega} + \frac{\kappa}{100}\mathbf{n}_{\text{noise}}\,\sigma(\mathbf{\Omega})\,, \tag{11}$$

where $\mathbf{\Omega}$ and $\hat{\mathbf{\Omega}}$ are respectively the noisefree and noisy feature matrices, $\sigma(\mathbf{\Omega})$ is a diagonal matrix whose elements are the standard deviations of the corresponding columns of $\mathbf{\Omega}$, κ is the noise level (percentage), and $\mathbf{n}_{\text{noise}}$ is a random independent variables matrix of the same size as $\mathbf{\Omega}$ following a standard normal distribution.

Figure 8 shows the histogram of the features with different percentages of noise.

[5] Learners that are slightly better than random guesses.

Next, we classify the noisy features using the three different MLAs identified in Section 5 as being the most effective, namely ensemble, SVM and KNN, in a 50% hold-out classification task. In other words, 50% of the data is used for training and the remaining 50% is used for testing, to ensure that the trained MLAs has not seen the test set before.

It is observed (Figure 9) that the average accuracy drops from 99.9% using noisefree features, to 91.5% when the features are contaminated by 10% noise, and further to 76.7% using features with 20% noise. All three classifiers had similar performance.

One should note however, that the 20% noise scenario is quite conservative. The reason is that noise is applied directly to the features here (i.e. to the centre frequencies of the decomposed modes), rather than to the raw signal. In reality the ultrasonic signals can contain noise due to different intervening factors in the testing regime, such as variations in amount of the gel applied to the surface of the wood, pressure applied to the transducer/receiver, and misalignment of transducer/receiver. Most of these sources of noise will be mitigated by a low-pass filter as mentioned in Section 3.

The results of this section therefore demonstrate that the proposed feature extraction and classification procedure can satisfactorily solve the classification problem.

Next, we propose to investigate the performance of single features alone in solving the three binary sub-classification problems that collectively are equivalent to the full classification task of this paper. Their performance is evaluated using the confusion matrices, obtained solving the classification problems using SVM, KNN and Ensemble methods.

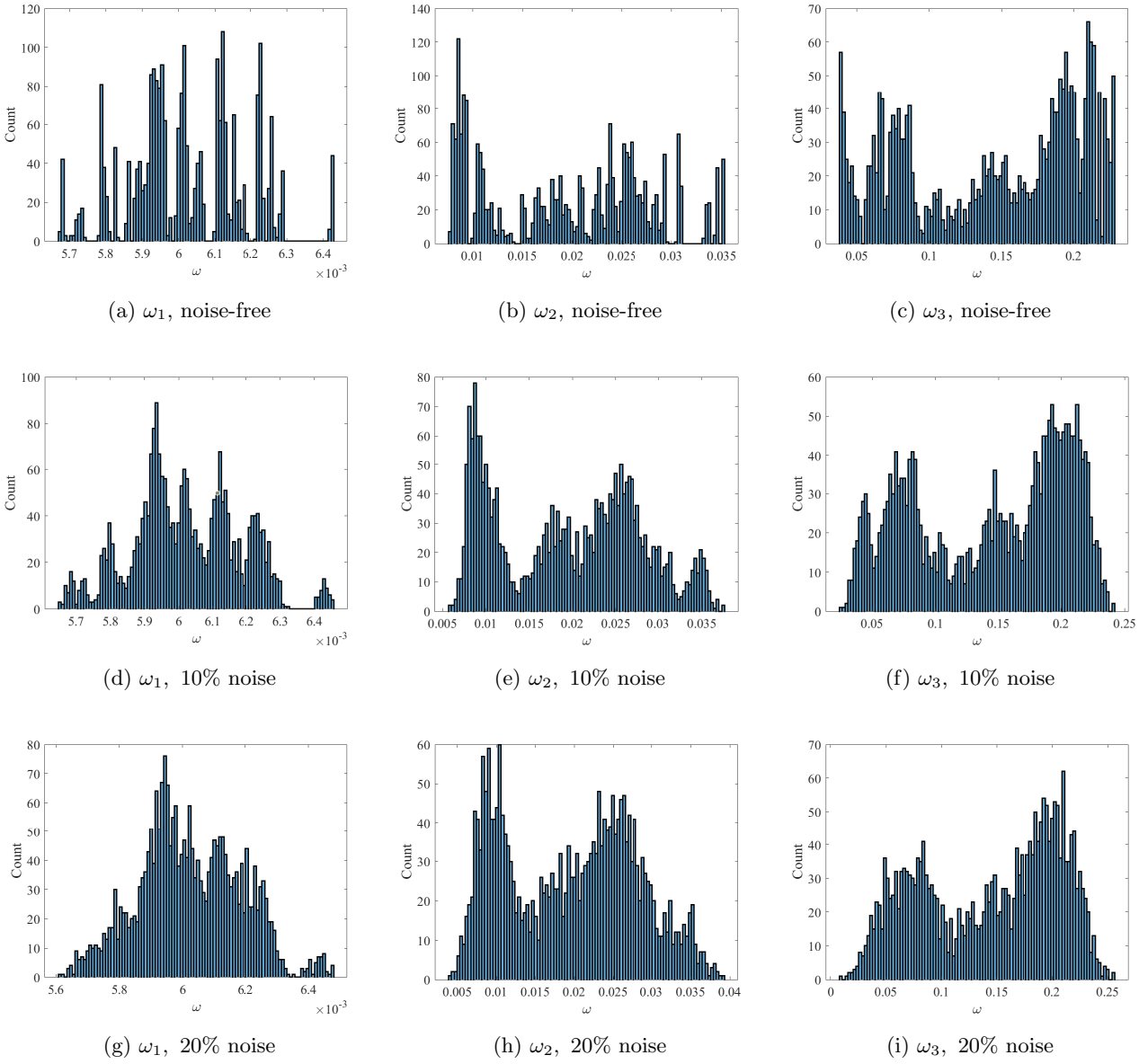
7. Sub-classification role of features

Assume a case where we use only one of the features in the classification problem and there are at least three features/decomposition modes available. Running a machine algorithm on the classification task using only that one feature, we aim to derive information from the confusion matrix about the ability of a single feature alone to recognise the subclasses in the dataset.

To this end, take $s = 1 \dots S$ to represent the sub-classification scenarios, where $S = 3$ in the problem of this paper, and $s = 1$, $s = 2$, and $s = 3$ indicate respectively the sub-classifications regarding the "type" (species) of wood, "damage" state of the wood, and "direction" of the test. In this exercise the classifications are binary, i.e. we do not distinguish between small and large defects. We obtain a value to characterise the relative false discovery rate (RFDR) regarding the s^{th} sub-classification problem using the confusion matrix as follows,

$$\text{RFDR}_s^{(m)} = \frac{\sum_{l=1}^{L} \left(a_l^{(s)} \times C_l^{(m)} \right)}{\sum_{s=1}^{S} \sum_{l=1}^{L} \left(a_l^{(s)} \times C_l^{(m)} \right)}; \tag{12}$$

where L represents the number of labels (12 in this paper), m and l are respectively the feature and label indices, $C_l^{(m)}$ indicates the l^{th} column of the confusion matrix, obtained using only the m^{th} feature in the classification problem, and $a_l^{(s)}$ is a binary vector indicating which test class correspond to negative



(a) ω_1, noise-free (b) ω_2, noise-free (c) ω_3, noise-free

(d) ω_1, 10% noise (e) ω_2, 10% noise (f) ω_3, 10% noise

(g) ω_1, 20% noise (h) ω_2, 20% noise (i) ω_3, 20% noise

Figure 8: Histogram of the three features with different percentage of noise.

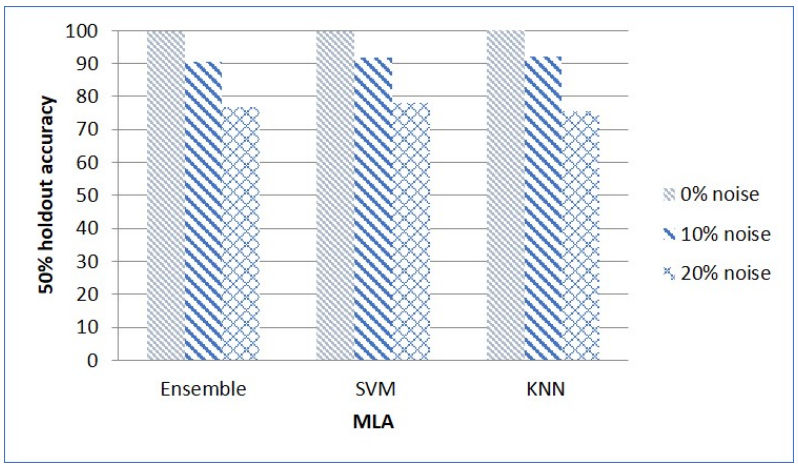


Figure 9: Percentage accuracy results of the optimised ensemble, SVM, and KNN algorithms in 50% hold-out classification using 10 and 20 percent of three noisy features.

result for a given label l and sub-classification s. Referring to Table 1, $a_l^{(s)}$ is obtained, for $s = 1, 2, 3$, as follows,

$$a_l^{(1)} = \begin{cases} [0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1]; & \text{for } \; l = 1, \ldots 6 \\ [1\,1\,1\,1\,1\,1\,0\,0\,0\,0\,0\,0]; & \text{for } \; l = 7, \ldots 12 \end{cases} \tag{13}$$

$$a_l^{(2)} = \begin{cases} [0\,1\,1\,0\,1\,1\,0\,1\,1\,0\,1\,1]; & \text{for } \; l = 1, 4, 7, 10 \\ [1\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0]; & \text{for } \; l = 2, 3, 5, 6, 8, 9, 11, 12 \end{cases} \tag{14}$$

$$a_l^{(3)} = \begin{cases} [0\,0\,0\,1\,1\,1\,0\,0\,0\,1\,1\,1]; & \text{for } \; l = 1, 2, 3, 7, 8, 9 \\ [1\,1\,1\,0\,0\,0\,1\,1\,1\,0\,0\,0]; & \text{for } \; l = 4, 5, 6, 10, 11, 12 \end{cases} \tag{15}$$

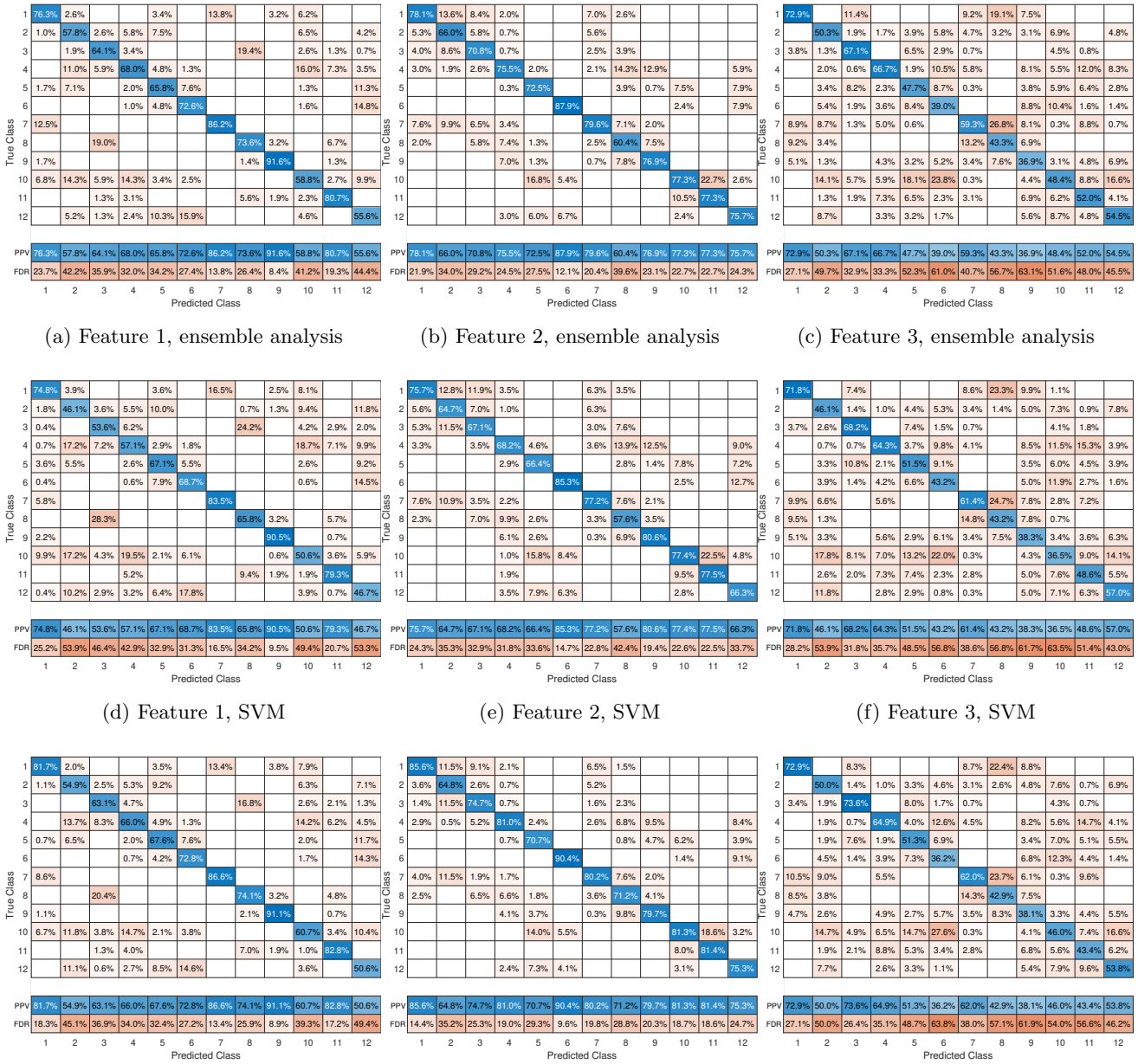
Recalling that S is the number of sub-classification cases, it is noted that for the m^{th} feature, it follows from Equation 12 that $\sum_s \text{RFDR}_s^{(m)} = 1$.

Next, the performance of the first three features is evaluated using the above strategy. To that end, the three MLAs identified above as being the most accurate, i.e. ensemble, SVM, and KNN models, are optimised using only one of the features obtained from 3 decomposition modes. Equation 12 is then used to evaluate the RFDR metric corresponding to the performance of each feature in the sub-classification problem. Since the three chosen MLAs all produce similar results, the RFDR metric is not expected to vary much from one machine learning model to another in describing the ability of the features to explore every sub-classification task.

Figure 10 shows the confusion matrices obtained respectively from the ensemble, SVM and KNN models with only one of the features being used in each case. The red shaded off-diagonal numbers are the rates of false positive detections (expressed as a percentage), with zero values indicating ideal performance. The columns are summed to give the overall False Discovery Rates (FDR) for each class, which are presented at the bottom of the figure for each case. The complement of the FDR is the overall Positive Predictive Value (PPV), indicated by the blue shaded number above the FDR. The PPV is also indicated on the diagonal of the matrix, representing the true positive detection rate, where values of 100% represent the ideal case.

Figure 11 shows the RFDR results for each sub-classification based on a single feature using the aforementioned machine learning models. Results from Features 1, 2 and 3 are shown in Figures 11a, 11b, and 11c, respectively. It is seen that very similar results are achieved using the three different MLAs. Note that the applied machine learning techniques follow different strategies to solve the classification problem. Therefore, we can conclude that the proposed RFDR metric is reliable, and we next discuss what information can be derived from the proposed RFDR metric.

The results for $\text{RFDR}_1^{(1)}$, $\text{RFDR}_1^{(2)}$, and $\text{RFDR}_1^{(3)}$ indicate that all the features when used alone perform relatively poorly at recognising the type of the wood relative to other sub-classification tasks. Note that good performance is associated with a low $\text{RFDR}_s^{(m)}$, so the performance of the features is ranked as Feature 3, Feature 2, and Feature 1. The poor performance is surprising as the material does



(a) Feature 1, ensemble analysis (b) Feature 2, ensemble analysis (c) Feature 3, ensemble analysis

(d) Feature 1, SVM (e) Feature 2, SVM (f) Feature 3, SVM

(g) Feature 1, KNN (h) Feature 2, KNN (i) Feature 3, KNN

Figure 10: Confusion matrices obtained from using different MLAs along with one of the features.

significantly affect the wave velocities, which in turn should affect component frequencies. However, it is noted that the wave velocities are not as different as one might expect, since the stiffness and density are correlated.[6] Furthermore, there is considerably greater variability with the direction than with the wood type, which, combined with the natural variability between nominally identical samples, may mask the overall impact of wood type. What is interesting though is that the different wood types have a different pattern of variation of acoustic velocity with direction, which may account for the considerably better results when classifying with multiple features.

[6]The wave velocity varies considerably, depending mainly on the direction of travel and to some extent on the wood type, approximately within the range 1300-2200 m/s in the radial and tangential directions for the specimens in this study [48].

The damage state of the wood, in contrast, is better able to be captured by a single feature, but the first feature does this more successfully than the other two, as seen comparing the results of $\mathrm{RFDR}_2^{(1)}$, $\mathrm{RFDR}_2^{(2)}$, and $\mathrm{RFDR}_2^{(3)}$. The first feature is the center frequency of the first IMF, which is the lowest of the component frequencies. Referring to Figure 4, this is around 6.3 kHz, and corresponds to radial and tangential half-wavelengths of roughly 100–175 mm based on known acoustic velocities.[7] These are very comparable to the specimen cross-section dimensions in the radial-tangential plane (90×90 mm^2), and it is known that acoustic standing wave patterns are modified by geometric features of size comparable to the wavelength. While the longitudinal wavelength is not known, it would be longer, hence again of comparable order to the specimen length (300 mm). The results therefore suggest that the lower frequency acoustic modes within the sample play a significant role in Feature 1. One might have expected this to correlate more strongly with wood type and grain, but this is not the case. Therefore, the results suggest that the presence of damage possibly significantly modifies the lower mode acoustic standing wave patterns within the sample, accounting for its impact on Feature 1. Further work is therefore suggested on different sized samples.

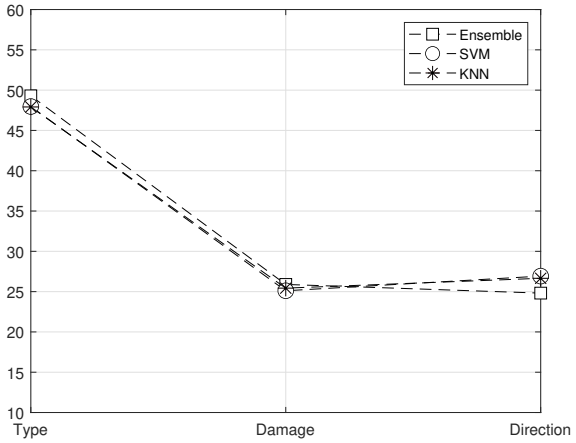
Results of $\mathrm{RFDR}_3^{(1)}$, $\mathrm{RFDR}_3^{(2)}$, and $\mathrm{RFDR}_3^{(3)}$ show that the second feature is relatively more successful in distinguishing the direction of the test, or equivalently, the direction of the wood grains, though all features do a reasonable job. It is observed in Figure 4 that Feature 2 is the most variable of the three, and (in this figure at least) the variability is correlated with the direction. This finding is therefore not surprising.

8. Field trial

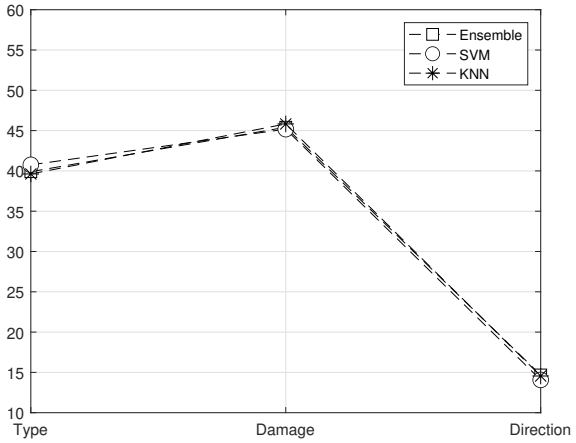
Self-pruning process in trees may well initiate natural imperfections in the trunk of the defected trees. Figure 12 shows examples of the trunk of trees with natural imperfections. The wood at the defected area is of lower quality and can be scraped out by an instrument. Therefore, it is vital to identify and avoid harvesting trees with these imperfections, as the trunk of these trees is not usually useful for industrial purposes.

In this section, the proposed features are further used to classify billets selected from standing trees harvested from the site Collie in WA, Australia [57]. The billets, of random geometry, were tested through different randomly selected angles, while debarked at the location upon which the transducer and receiver were placed. All the tested billets were assigned a label either as healthy or defective through visual inspection after cutting. Table 3 shows the meteorological conditions at the site at the time of testing. Table 4 lists the number of healthy and defective specimens and the overall number of the tests conducted. Note that the billets were tested through the radial direction with respect to the growth rings of wood. Therefore, this section aims to classify based on the health condition of the specimens only.

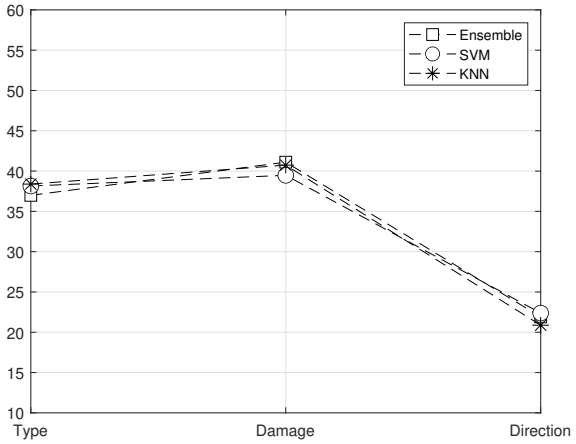
[7] See previous note.



(a) Feature 1



(b) Feature 2



(c) Feature 3

Figure 11: RFDR obtained from different MLAs in percentage for the sub-classification tasks using only one of the features each time.

The best 5-fold cross-validation accuracy result of 88.6% was obtained from the Ensembles method, using center frequencies of three decompositions via VMD, as shown in Figure 13a. This obviously

(a) Original (b) Scraped out

Figure 12: Examples of bulk of standing trees with natural imperfections. The wood at the defected area is soft in quality and can be scraped out by an instrument.

Table 3: The meteorological condition upon testing at the site Collie in WA, Australia.

State	**Site**	**Wood species**	**Temperature (°C)**	**Humidity (%)**
WA	Collie	Jarrah	5.1	90

indicate the more challenging nature of the problem compared with the classification problem of the laboratory trial. Obviously, exploiting more descriptive features would be beneficial in this case. As such, the Root Mean Square of the Instantaneous Frequency (IF) signal, shown as $\mathrm{RMS_{IF}}$ pertaining to each IMF is also extracted and used in the training process [46]. The $\mathrm{RMS_{IF}}$ for an IMF is obtained as follows:

$$\mathrm{RMS_{IF}} = \sqrt{\frac{\sum_{i=1}^{n} \omega(t)^2}{n}}, \tag{16}$$

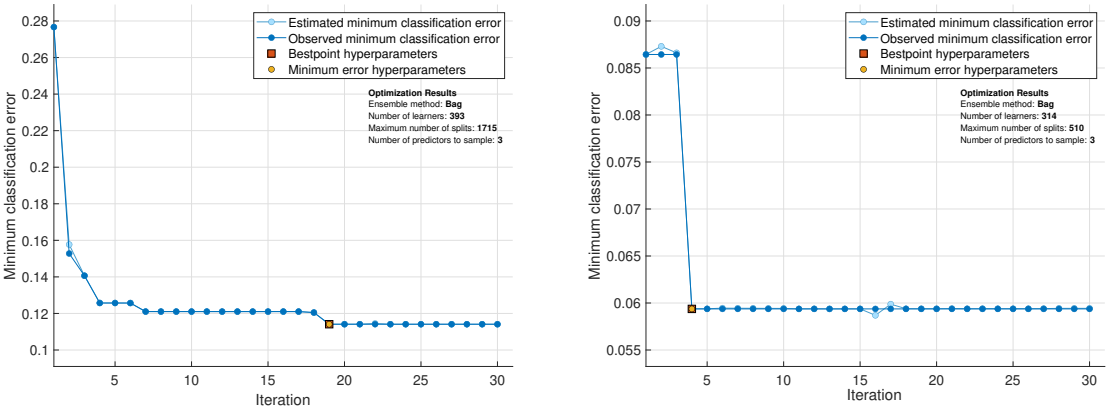
where $\omega(t)$ is the instantaneous frequency of the IMF, and is obtained from the Gabor's analytical signal $u_a(t) = u(t) + j\hat{u}(t)$, whose real and imaginary parts are respectively the original IMF and its Hilbert transform [58]. As such one can obtain $\omega(t)$ as follows:

$$\omega(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\tan^{-1}\left(\frac{\hat{u(t)}}{u(t)}\right)\right), \tag{17}$$

Using the obtained $\mathrm{RMS_{IF}}$ features along with the center frequency features obtained from three IMFs (six features in total), the classification problem was solved again. This time 94.2% 5-fold cross-validation accuracy was achieved, using the optimised ensemble of classifiers, as shown in Figure 13b. The results indeed indicate the effectiveness of the proposed approach for health monitoring of standing trees.

Table 4: The number of billets and ultrasonic tests conducted on billets harvested from the site Collie in WA, Australia.

Number of billets		
Condition	**Billets #**	**Ultrasonic tests #**
Intact	37	838
Defective	37	897



(a) Center frequency features (88.6%)

(b) Center frequency & RMS_{IF} features (94.2%)

Figure 13: 5-fold cross-validation accuracy results obtained from the optimised Ensemble of classifiers using different features.

9. Conclusions and future work

This paper solves the problem of classifying wood type, condition and grain direction, using MLAs applied to features extracted using VMD from ultrasonic signals obtained from CU tests conducted on two types of woods.

To this end, the contact ultrasonic device is proposed as an effective tool for collecting data from wood specimens, and the procedure of the proposed machine learning strategy based on the test results is depicted in Figure 14. The ultrasonic signals were first decomposed into their nonlinear oscillatory modes using VMD, an advanced signal decomposition algorithm. VMD is a parametric decomposition algorithm, and it is crucial to choose its parameters properly, which was fully discussed.

A key parameter is the number of IMFs K into which the signal is to be decomposed. The center frequencies of the modes were taken as features, therefore k also defines the number of features. We showed, by calculating the percentage contribution from each feature to principal components (Equation 9), that different numbers of decomposition modes K do not compromise the solution of the classification problem. We further argued that the features contributing to the first principal component are the most important as they contributed most variability to the dataset. However, it was established that three features/decomposition modes is sufficient and effective when using either ensemble, SVM, or KNN algorithms.

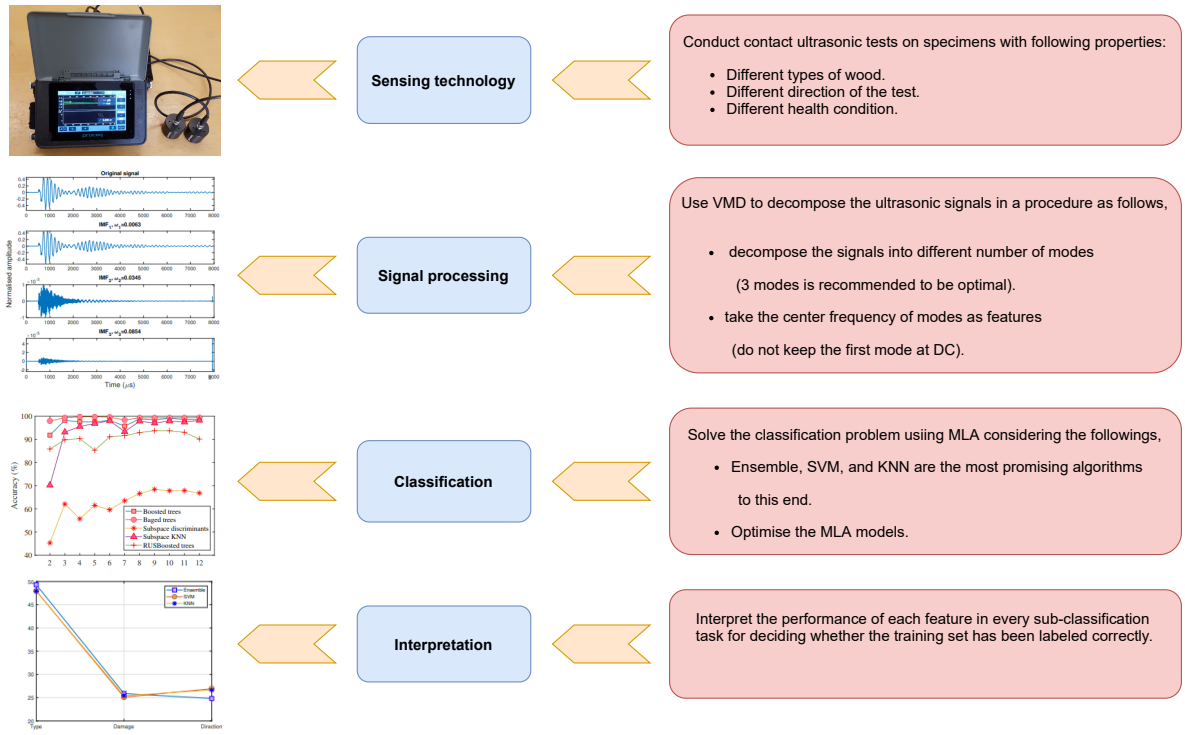


Figure 14: The procedure of classification of wood properties using the proposed strategy in this paper.

To ensure that the proposed strategy can properly solve the classification problem in the presence of significant noise, the extracted features were contaminated with 10% and 20% noise levels. Greater than 90% and 75% hold-out accuracy was achieved respectively with these noise levels. However, in practice these noise levels can be avoided by using a low-pass filter prior to the decomposition. We further optimised each model hyperparameters using a Bayesian optimisation algorithm embedded in the MATLAB machine learning toolbox.

In Section 7, we proposed the metric RFDR to explore the contribution of each feature to the different sub-classification tasks embedded in the main classification problem. This involved solving the classification problem using a single feature, and using the confusion matrix to calculate the false discovery rate associated with each sub-classification task. We showed that the proposed metric informs how each feature explores different aspect of the classification problem (i.e. the type of wood, the state of damage, and direction of the tests), and is independent of the MLA used to solve the problem.

The strategy followed for solving the classification problem of this paper provides a strong tool to explore how features should be derived from a nonlinear signal. Moreover, the proposed RFDR metric was shown to be informative in terms understanding aspects of the physics of the ultrasonic waves and their interactions with attributes of the samples that we sought to classify. The results of this paper, therefore, acknowledge that although MLAs are in some senses 'black boxes', their results (such as by using the confusion matrix in this paper) can be interpreted to explore the physics of problems.

The proposed strategy in this paper can be used in many applications, including damage detection of wooden poles, quality assessment of standing trees, and damage identification and classification of

historical wooden heritage structures.

Some variables in contact ultrasonic tests are difficult to control or to measure, such as amount of the gel applied to the surfaces, pressure applied to the transducer and receiver, and any misalignment of the transducer and receiver. These were not controlled for in the present work, therefore it can be concluded that the MLA was able to successfully classify sample in spite of variability of these factors.

On the other hand, in spite of performing 2,400 tests, only a single sample geometry was considered in the lab trial section. Therefore, in order to test the effectiveness of the proposed features extracted from the time-frequency domain of ultrasonic signals in real world problems, billets with and without natural imperfections harvested from the site Collie in WA, Australia, waere classified. A 94.2% 5-fold cross-validation accuracy was obtained from optimised ensemble of classifiers when solving the problem using the center frequency features along with the $\mathrm{RMS_{IF}}$ features, based on [46], demonstrating the effectiveness of the proposed features in classifying wood materials based on their health condition.

References

[1] H. Yang, L. Yu, Feature extraction of wood-hole defects using wavelet-based ultrasonic testing, Journal of forestry research 28 (2) (2017) 395–402.

[2] W. Li, J. Van den Bulcke, D. Mannes, E. Lehmann, I. De Windt, M. Dierick, J. Van Acker, Impact of internal structure on water-resistance of plywood studied using neutron radiography and X-ray tomography, Construction and Building Materials 73 (2014) 171–179.

[3] G. López, L. A. Basterra, L. Acuña, Estimation of wood density using infrared thermography, Construction and Building Materials 42 (2013) 29–32.

[4] T. Goto, Y. Tomikawa, S. Nakayama, T. Furuno, Changes of propagation velocity of ultrasonic waves and partial compression strength of decay-treated woods relationship between decrease of propagation velocity of ultrasonic waves and remaining strength, Mokuzai Gakkaishi 57 (6) (2011) 359–369.

[5] S. Lee, S.-J. Lee, J. S. Lee, K.-B. Kim, J.-J. Lee, H. Yeo, Basic study on nondestructive evaluation of artificial deterioration of a wooden rafter by ultrasonic measurement, Journal of Wood Science 57 (5) (2011) 387–394.

[6] F. Tallavo, G. Cascante, M. D. Pandey, A novel methodology for condition assessment of wood poles using ultrasonic testing, NDT & E International 52 (2012) 149–156.

[7] T. Mori, Y. Yanase, K. Tanaka, K. Kawano, Y. Noda, M. Mori, H. Kurisaki, K. Komatsu, Evaluation of compression and bending strength properties of wood damaged from bio-deterioration, Journal of the Society of Materials Science, Japan 62 (4) (2013) 280–285.

[8] S.-J. Lee, S. Lee, S.-J. Pang, C.-K. Kim, K.-M. Kim, K.-B. Kim, J.-J. Lee, Indirect detection of internal defects in wooden rafter with ultrasound, Journal of the Korean Wood Science and Technology 41 (2) (2013) 164–172.

[9] A. Ettelaei, M. Layeghi, H. Z. Hosseinabadi, G. Ebrahimi, Prediction of modulus of elasticity of poplar wood using ultrasonic technique by applying empirical correction factors, Measurement 135 (2019) 392–399.

[10] E. Blomme, D. Bulcaen, F. Declercq, Air-coupled ultrasonic nde: experiments in the frequency range 750 kHz–2 MHz, NDT & E International 35 (7) (2002) 417–426.

[11] A. C. Senalik, G. Schueneman, R. J. Ross, Ultrasonic-based nondestructive evaluation methods for wood: a primer and historical review, USDA Forest Service, Forest Products Laboratory, General Technical Report, FPL-GTR-235, 2014; 36 p. 235 (2014) 1–36.

[12] W. H. Emmingham, S. A. Fitzgerald, et al., Pruning to enhance tree and stand value (1995).

[13] M. Mori, M. Hasegawa, J.-C. Yoo, S.-G. Kang, J. Matsumura, Nondestructive evaluation of bending strength of wood with artificial holes by employing air-coupled ultrasonics, Construction and Building Materials 110 (2016) 24–31.

[14] T. Palander, J. Eronen, K. Kärhä, H. Ovaskainen, Development of a wood damage monitoring system for mechanized harvesting, Annals of Forest Research 61 (2) (2018) 243–258.

[15] U. Dackermann, R. Elsener, J. Li, K. Crews, A comparative study of using static and ultrasonic material testing methods to determine the anisotropic material properties of wood, Construction and Building Materials 102 (2016) 963–976.

[16] S. Bandara, P. Rajeev, E. Gad, B. Sriskantharajah, I. Flatley, Damage detection of in service timber poles using hilbert-huang transform, NDT & E International 107 (2019) 102141.

[17] B. Sriskantharajah, E. Gad, S. Bandara, P. Rajeev, I. Flatley, Condition assessment tool for timber utility poles using stress wave propagation technique, Nondestructive Testing and Evaluation 36 (3) (2021) 336–356 .

[18] B. İpekoğlu, An architectural evaluation method for conservation of traditional dwellings, Building and environment 41 (3) (2006) 386–394.

[19] A. Salmi, K. Steffen, J. Eskelinen, E. Hæggström, Ultrasonic quantitative strength assessment of artificially aged and archaeological wood samples, The Journal of the Acoustical Society of America 123 (5) (2008) 3607–3607.

[20] H. Cruz, D. Yeomans, E. Tsakanika, N. Macchioni, A. Jorissen, M. Touza, M. Mannucci, P. B. Lourenço, Guidelines for on-site assessment of historic timber structures, International Journal of Architectural Heritage 9 (3) (2015) 277–289.

[21] M. M. Conde, C. R. Liñán, P. R. de Hita, Use of ultrasound as a nondestructive evaluation technique for sustainable interventions on wooden structures, Building and Environment 82 (2014) 247–257.

[22] M. Hirao, H. Ogi, EMATs for science and industry: noncontacting ultrasonic measurements, Springer Science & Business Media, 2013.

[23] L. Drain, Laser ultrasonics techniques and applications, Routledge, 2019.

[24] W. Grandia, C. Fortunko, Nde applications of air-coupled ultrasonic transducers, in: 1995 IEEE Ultrasonics Symposium. Proceedings. An International Symposium, Vol. 1, IEEE, 1995, pp. 697–709.

[25] Y. Fang, L. Lin, H. Feng, Z. Lu, G. W. Emms, Review of the use of air-coupled ultrasonic technologies for nondestructive testing of wood and wood products, Computers and electronics in agriculture 137 (2017) 79–87.

[26] D. Chimenti, Review of air-coupled ultrasonic materials characterization, Ultrasonics 54 (7) (2014) 1804–1816.

[27] T. Marhenke, J. Neuenschwander, R. Furrer, J. Twiefel, J. Hasener, P. Niemz, S. J. Sanabria, Modeling of delamination detection utilizing air-coupled ultrasound in wood-based composites, NDT & E International 99 (2018) 1–12.

[28] N. Jain, Laser ultrasonics: The next big nondestructive inspection technology?, Quality magazine, August 3, 2011. Access date: November 6, 2020.
URL https://www.qualitymag.com/articles/90017

[29] S. K. Pedram, P. Mudge, T.-H. Gan, Enhancement of ultrasonic guided wave signals using a split-spectrum processing method, Applied sciences 8 (10) (2018) 1815.

[30] J. El Najjar, S. Mustapha, Condition assessment of timber utility poles using ultrasonic guided waves, Construction and Building Materials 272 (2021) 121902.

[31] U. Dackermann, B. Skinner, J. Li, Guided wave–based condition assessment of in situ timber utility poles using machine learning algorithms, Structural Health Monitoring 13 (4) (2014) 374–388.

[32] B. Marchetti, R. Munaretto, G. Revel, E. P. Tomasini, V. B. Bianche, Non-contact ultrasonic sensor for density measurement and defect detection on wood, in: Proc. 16th World Conference on Nondestructive Testing, 2004, pp. 14–21.

[33] H. Yanagida, Y. Tamura, K.-M. Kim, J. J. Lee, Development of ultrasonic time-of-flight computed tomography for hard wood with anisotropic acoustic property, Japanese Journal of Applied Physics 46 (8R) (2007) 5321.

[34] N. Yaitskova, J. W. van de Kuilen, Time-of-flight modeling of transversal ultrasonic scan of wood, The Journal of the Acoustical Society of America 135 (6) (2014) 3409–3415.

[35] G. Karaiskos, A. Deraemaeker, D. Aggelis, D. Van Hemelrijck, Monitoring of concrete structures using the ultrasonic pulse velocity method, Smart Materials and Structures 24 (11) (2015) 113001.

[36] D. W. Green, J. E. Winandy, D. E. Kretschmann, Mechanical properties of wood. Wood handbook: wood as an engineering material (Chapter 4), Vol. GTR-113, Madison, WI : USDA Forest Service, 1999.

[37] P. Mackenzie-Helnwein, J. Eberhardsteiner, H. A. Mang, A multi-surface plasticity model for clear wood and its application to the finite element analysis of structural details, Computational Mechanics 31 (1-2) (2003) 204–218.

[38] J. Liu, J. Zhou, S. Qi, Z. Cui, K. Wang, Influence of texture anisotropy on acoustoelastic birefringence in stressed wood, The Journal of the Acoustical Society of America 134 (5) (2013) 4227–4227.

[39] EN 408:2010+A1:2012 – Timber structures, Structural timber and glued laminated timber, Determination of some physical and mechanical properties, Standard, European Standards (2010).

[40] M. H. Ramage, H. Burridge, M. Busse-Wicher, G. Fereday, T. Reynolds, D. U. Shah, G. Wu, L. Yu, P. Fleming, D. Densley-Tingley, et al., The wood from the trees: The use of timber in construction, Renewable and Sustainable Energy Reviews 68 (2017) 333–359.

[41] C. . Handbook, Wood, Wood as an engineering material. forest products laboratory, Gen. Tech. Rep. FPL–GTR–113, USDA Product Society, Madison, Wisconsin, USA (1999).

[42] S. Lee, S.-J. Lee, J. S. Lee, K.-B. Kim, J.-J. Lee, H. Yeo, Basic study on nondestructive evaluation of artificial deterioration of a wooden rafter by ultrasonic measurement, Journal of wood science 57 (5) (2011) 387–394.

[43] L. Reinprecht, M. Pánek, Ultrasonic technique for evaluation of bio-defects in wood: Part 1–influence of the position, extent and degree of internal artificial rots, International Wood Products Journal 3 (2) (2012) 107–115.

[44] Y. Tian, Y. Shen, D. Rao, W. Xu, Metamaterial improved nonlinear ultrasonics for fatigue damage detection, Smart Materials and Structures 28 (7) (2019) 075038.

[45] H. Liu, S. Guo, Y. F. Chen, C. Y. Tan, L. Zhang, Acoustic shearography for crack detection in metallic plates, Smart Materials and Structures 27 (8) (2018) 085018.

[46] M. Mousavi, A. H. Gandomi, Wood hole-damage detection and classification via contact ultrasonic testing, Construction and Building Materials 307 (2021) 124999.

[47] V. Nasir, H. Fathi, S. Kazemirad, Combined machine learning–wave propagation approach for monitoring timber mechanical properties under UV aging, Structural Health Monitoring 20 (4) (2021) 2035–2053.

[48] M. Mousavi, M. S. Taskhiri, D. Holloway, J. Olivier, P. Turner, Feature extraction of wood-hole defects using empirical mode decomposition of ultrasonic signals, NDT & E International (2020) 102282.

[49] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Transactions on Signal Processing 62 (3) (2014) 531–544.

[50] D. Zosso, Variational mode decomposition, matlab central file exchange (Retrieved August 27, 2020).
URL https://www.mathworks.com/matlabcentral/fileexchange/44765-variational-mode-decomposition

[51] Y. Wang, R. Markert, J. Xiang, W. Zheng, Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system, Mechanical Systems and Signal Processing 60 (2015) 243–251.

[52] H. Abdi, L. J. Williams, Principal component analysis, Wiley interdisciplinary reviews: computational statistics 2 (4) (2010) 433–459.

[53] Y.-S. Shih, Families of splitting criteria for classification trees, Statistics and Computing 9 (4) (1999) 309–315.

[54] L. Rokach, O. Z. Maimon, Data mining with decision trees: theory and applications, Vol. 69, p. 57, World scientific, 2008.

[55] C. M. Biship, Pattern recognition and machine learning (information science and statistics) (2007, P 183.).

[56] Z.-H. Zhou, Ensemble learning., Encyclopedia of biometrics 1 (2009) 270–273.

[57] M. Mousavi, M. S. Taskhiri, A. H. Gandomi, Mechanised harvesting of standing trees, under review.

[58] M. Mousavi, D. Holloway, J. Olivier, A. H. Gandomi, Beam damage detection using synchronisation of peaks in instantaneous frequency and amplitude of vibration data, Measurement 168 (2021) 108297.