A Novel Feature Selection Framework to Deal with Outliers in SHM Problems

Mohsen Mousavi[a], Amir H. Gandomi *[a,b,c]

[a] *Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW 2007, Australia*
[b] *University Research and Innovation Center (EKIK), Óbuda University, 1034 Budapest, Hungary*
[c] *NASK National Research Institute, ul. Kolska 12, Warsaw, 01-045, Poland*

Abstract

Predictive modeling becomes increasingly reliant on feature selection algorithms as dataset dimensions grow. Feature selection methods can estimate biased feature importance values when datasets present outliers. This can compromise the effectiveness of classification algorithms by reducing their accuracy. This paper primarily targets feature selection for solving damage classification problems in civil and mechanical engineering. We propose a novel framework for feature selection based on the minimum covariance determinant and PCA. The proposed feature selection algorithm obtains a metric for scoring features' importance based on the loading matrix obtained from the robust PCA algorithm applied to the feature matrix. The covariance matrix of the robust PCA algorithm is obtained from the minimum covariance determinant algorithm. This way, features considered outliers in the feature matrix are discarded from further analysis. The proposed feature extraction framework is tested on several damage classification problems of wood materials. Its superiority is demonstrated by comparing its results with a PCA-based feature selection algorithm. The results obtained from the proposed unsupervised feature selection method demonstrate its robustness to outliers, rendering it a viable application technique in complex problems involving datasets containing outliers.

Keywords: `Civil engineering, Damage classification, Feature importance, SHM`

Email addresses: `mohsen.mousavi@uts.edu.au` (Mohsen Mousavi), `gandomi@uts.edu.au` (Amir H. Gandomi *)

1. Introduction

Various fields have generated exponentially more data in recent years. This data, which is often high-dimensional, contains a large number of features that can sometimes be irrelevant, redundant, or noisy. As such, feature selection techniques are often utilized to extract the most relevant features from a dataset to enhance machine learning model performance, reduce overfitting, and increase interpretability. Principal Component Analysis (PCA) is one of the popular techniques for feature selection in various fields. In this paper, we propose a robust feature selection technique, an improved version of our recently proposed approach in [1], that overcomes some limitations of existing methods.

Using PCA, one can reduce the dimensionality of high-dimensional data while retaining as much variance as possible. A PCA-based feature selection technique aims to extract the most relevant features from a dataset through PCA analysis of the feature matrix. The most commonly used PCA-driven feature selection methods are subset selection and ranking based on PCA.

PCA-based feature selection methods are widely used to assign scores to features based on the eigenvalues of principal components, with higher scores indicating more importance and selection for further analysis. Classical methods include the correlation-based feature selection (CFS) algorithm proposed by Hall [2] and the Relief algorithm proposed by Kira [3]. CFS selects features with a high correlation to the class label and low correlation to other features, while Relief selects features based on their ability to distinguish between instances of different classes. Existing feature selection methods have limitations in handling nonlinear and highly correlated features and dealing with multicollinearity and noisy features [4, 5].

Despite their widespread use, PCA-based feature selection techniques have several limitations, including sensitivity to outliers, assumptions of Gaussian distribution in the data, and difficulties in handling nonlinear and highly correlated features. These limitations call for more robust approaches to PCA-based feature selection that can overcome these challenges [6].

In this paper, we propose a novel and robust PCA-based feature selection technique that addresses the limitations of existing methods. Our proposed approach incorporates a modified version of our previously proposed PCA-based feature selection method that can handle nonlinear and highly correlated features, multicollinearity, and noisy features. The proposed method is unsupervised in the sense that it does not rely on the features' labels. The introduced robust measure of variance, based on the concept of Minimum Covariance Determinant (MCD), makes the proposed method less sensitive to outliers.

2. Case study

This paper considers the classification problem of standing trees based on their health condition originally presented in [1]. Several billets harvested from various sites in Australian states, i.e., New South Wales (NSW) and Western Australia (WA), were tested using an ultrasound device, i.e., Pundit

Figure 1: An example of a defective wood trunk.

Table 1: Type of studied wood at different sites and the meteorological conditions upon testing [1].

State	Site	Wood species	Temperature (°C)	Humidity (%)
WA	Collie	Jarrah	5.1	90
NSW	Coffs harbour	Blackbutt & Greygum	10	90

PL-200. An example of a defective wood trunk is shown in Figure 1. Table 1 summarizes information about the type of wood and the environmental conditions at different sites upon testing.

The specimens were examined in various randomly chosen orientations to conduct the tests. After visually inspecting the billets, two various labels were assigned to the test results, which were healthy and defective. Table 2 presents data on the number of billets obtained from various locations and the total number of ultrasonic tests performed.

The ultrasonic test results were processed using the VMD algorithm to extract the necessary features for training the machine learning models (MLA). The feature extraction procedure is discussed in Section 3.1.

Table 2: The number of billets and the corresponding ultrasonic tests performed on wood samples at various sites, as documented in [1].

Number of billets		
Condition	**WA** #	**NSW** #
Intact	37	7
Defective	37	28
Number of ultrasonic test		
Condition	**WA** #	**NSW** #
Intact	838	213
Defective	897	617

3. Methodology

3.1. Feature extraction using VMD

The ultrasonic signal $S(t)$ was initially normalized by subtracting the mean and scaling by the range as follows [7, 8, 9]:

$$\overline{\mathbf{S}}(t) = \frac{\mathbf{S}(t) - \mu}{\max(\mathbf{S}(t)) - \min(\mathbf{S}(t))}. \tag{1}$$

where $S(t)$ and $\overline{S}(t)$ represent original signal and its mean-normalized version. A low-pass filter was then utilized to cut off the frequency content in each signal capped at 300 kHz [8]. VMD was employed to decompose the resulting signals to derive useful features. The theory of VMD is briefly outlined in this section as follows.

VMD has the following characteristics:

- It solves a variational optimization problem
- As a result, it decomposes a nonlinear/non-stationary signal into its Intrinsic Mode Functions termed IMF
- IMFs extracted through VMD are narrow-band signals and thus delineate the original signal's modes of oscillation.

The k^{th} IMF ($\mathbf{u}_k(t)$) of the VMD outcome is in the form of an amplitude/frequency modulated cosine signal as follows:

$$\mathbf{u}_k(t) = \mathbf{A}_k(t)\ \cos(\phi_k(t)), \tag{2}$$

where $\mathbf{A}_k(t)$ and $\phi_k(t)$ denote the instantaneous amplitude and phase, respectively. The instantaneous phase can be differentiated to derive the Instantaneous Frequency (IF) of the IMF as $\omega(t) = \frac{\partial \phi(t)}{\partial t}$. The IF can be more rigorously obtained by constructing Gabor's analytic signal ($\mathbf{u}_a(t)$) as follows [10]:

$$\mathbf{u}_a(t) = \mathbf{u}(t) + j\hat{\mathbf{u}}(t), \tag{3}$$

Table 3: The selected VMD parameters for the current problem [1].

Parameters	**Description**	**Specified values**
p	Number of IMFs	3
α	Denoising factor	N.A.
τ	Time interval	0.1
ϵ	Convergence threshold	10^{-5}
init	Center frequency initialiser	0
DC	Boolean parameter	0

in which, j and $\hat{\mathbf{u}}(t)$ represnt the imaginary unit and Hilbert transformation [11] of the given IMF signal $\mathbf{u}(t)$, respectively. As a result, one can work out the instantaneous frequency as follows:

$$\omega(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\tan^{-1}\left(\frac{\hat{\mathbf{u}(t)}}{\mathbf{u}(t)}\right)\right), \tag{4}$$

The following Lagrangian equation is optimized to update for the IMFs $\mathbf{u}_k$ and their corresponding center frequencies ω_k [12]:

$$\mathcal{L}(\mathbf{u}_k, \omega_k, \lambda) = \alpha \sum_k \left\| \partial_t \left(\delta(t) + \frac{j}{\pi t} * \mathbf{u}_k(t) \right) \times e^{-j\omega_k t} \right\|_2^2 + \left\| \mathbf{f}(t) - \sum_k \mathbf{u}_k(t) \right\|_2^2 + \left\langle \lambda(t), \mathbf{f}(t) - \sum_k \mathbf{u}_k(t) \right\rangle \tag{5}$$

VMD is a parametric decomposition algorithm requiring pre-specification of some parameters in its settings prior to applying the algorithm to a signal [13]. The parameters and their corresponding values used in the VMD algorithm are presented in Table 3. For more detailed information on how to specify the parameters for the VMD algorithm, readers are encouraged to refer to the works of Mousavi et al. [1, 8].

The IF signal corresponding to the i^{th} IMF was first constructed, and seven statistical features were then extracted as follows [1]:

1. The center frequency (ω) of the IMF_i, represented as x_{1i}.

2. The Root Mean Square (RMS) of the IF signal corresponding to the IMF_i signal, shown as x_{2i}, obtained as follows [8]:

$$\text{RMS}_{\text{IF}_i} = \sqrt{\frac{\sum_{i=1}^{n} \omega_i(t)^2}{n}}, \tag{6}$$

   where n is the length of the IMF_i, which is equal to the size of the original signal.

3. The first ($Q1_{\text{IF}}$), second ($Q2_{\text{IF}}$), and third ($Q3_{\text{IF}}$) quarterlies of the IF signal corresponding to the i^{th} IMF, shown as x_{3i}, x_{4i}, and x_{5i}, respectively.

4. The variance of the i^{th} IMF's IF signal (σ_{IF}), shown as x_{6i}.

Table 4: The description of all features naming.

Features	**Description**
x_{1i}	ω of IMF_i
x_{2i}	RMS_{IF} of IMF_i
x_{3i}	$Q1_{\text{IF}}$ of IMF_i
x_{4i}	$Q2_{\text{IF}}$ of IMF_i
x_{5i}	$Q3_{\text{IF}}$ of IMF_i
x_{6i}	k_{IF} of IMF_i
x_{7i}	σ_{IF} of IMF_i

5. The Kurtosis of the i^{th} IMF's IF signal (k_{IF}), shown as x_{7i}.

The extracted features derived from the IMFs of each signal are named as listed in Table 4.

3.2. Feature selection algorithm

Feature selection algorithms are designed to identify the most relevant and informative features for training purposes. This practice will ensure avoiding correlated features to make the process less time-consuming. It may also prevent over-fitting on the training set, decreasing the variance between the accuracy metrics obtained for the test and training sets.

Principal component analysis (PCA) is the favorite technique to rank features based on their importance to introduce variability in the feature space. As such, a feature's importance directly corresponds to its correlation with lower-order principal components (PCs). Consequently, features that exhibit a lower correlation with the lower-order PCs may have limited capability in describing the variability observed in the dataset. More details about obtaining each feature's correlation with different PCs are outlined below.

From now on, we will work with the standardized form of the feature matrix $\mathbf{X}$ obtained by centering with respect to its mean value and scaling by dividing by its standard deviation. The matrix $\mathbf{X}_{m\times p}$ contains features stacked in its columns with different observations occurring in rows, where it is assumed to be of rank $r \leq \min\{m, p\}$. Let's contemplate the singular value decomposition of matrix $\mathbf{X}$ in the following manner:

$$\mathbf{X} = \mathbf{P}\Delta\mathbf{Q}^{\text{T}} , \tag{7}$$

Here, we have matrices $\mathbf{P}_{m\times r}$ and $\mathbf{Q}_{p\times r}$, which represent the left and right singular vectors, respectively. It's worth mentioning that $\mathbf{Q}$ is a unitary matrix, meaning that its inverse is equal to its transpose, denoted as $\mathbf{Q}^{-1} = \mathbf{Q}^{\text{T}}$. Additionally, we have $\Delta_{r\times r}$, which represents the diagonal matrix of singular values.

The factor matrix $\mathbf{F}$ is defined as follows:

$$\mathbf{F} = \mathbf{P}\Delta , \tag{8}$$

whose columns represent the principal components of $\mathbf{X}$. Since $\mathbf{Q}$ is a unitary matrix, one can express (8) as:

$$\mathbf{F} = \mathbf{P}\Delta = \mathbf{P}\Delta\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{X}\,\mathbf{Q}\,. \tag{9}$$

Hence, $\mathbf{Q}$ can also be seen as a mapping matrix that projects the original feature matrix $\mathbf{X}$ onto the factor matrix $\mathbf{F}$. The "loading" matrix, representing the contribution of each feature to the principal components, is obtained by squaring the entries of $\mathbf{Q}$.

However, there can be events when the correlation between a feature with the second PC is much higher than another feature whose correlation is relatively more significant with the first PC.

3.3. Identifying challenges

The loading matrix, which was evaluated for the feature matrix of various sites, is depicted in Figure 2. The plots reveal that it can be challenging to objectively determine the relative importance of features in cases where, for example, the contribution of component x to PC1 is only marginally higher than that of feature y while the contribution of y to PC2 is significantly greater than that of x. To provide further clarity, let's consider features x_{11} and x_{12} in the chart shown in Figure 2a. The chart shows that the contribution of x_{12} to PC1 is 4.965%, which is higher than the contribution of x_{11} at 2.296%. However, the contribution of x_{11} to PC2 is 4.607%, significantly surpassing the contribution of x_{12} at 0.988%. This subjective variability in ranking feature importance necessitates the development of a robust feature selection algorithm to address this issue.

Next, the two-dimensional landscape of features extracted from the ultrasound tests is illustrated by plotting PC1 versus PC2, as shown in Figure 3. Figure 4 displays the box plot of the standardized VMD-extracted features. It is evident from the plot that several events in the dataset occur as outliers. Therefore, it is crucial to propose a robust feature selection algorithm that can effectively handle the presence of outliers.

Although our previously proposed technique in [1] addresses the first problem discussed above, it falls short in dealing with outliers in the feature space, as depicted in Figure 4. Hence, it is imperative to develop a robust algorithm to effectively address this limitation. In the subsequent section, we present a comprehensive overview of our previously proposed PCA-based method. Subsequently, we leverage the concept of the Minimum Covariance Determinant (MCD) to devise a modified, robust algorithm for feature selection. Finally, we systematically evaluate and compare the performance of the proposed MCD-based algorithm with the PCA-based method, utilizing various Machine Learning Algorithms (MLAs) in the context of solving the classification problem pertaining to wood.

4. A robust feature selection method

The methodology proposed in this paper builds upon the authors' previous work in [1]. Before discussing the new method, let's first review the summarized version of the proposed technique in [1], which serves as the foundation for the current research:

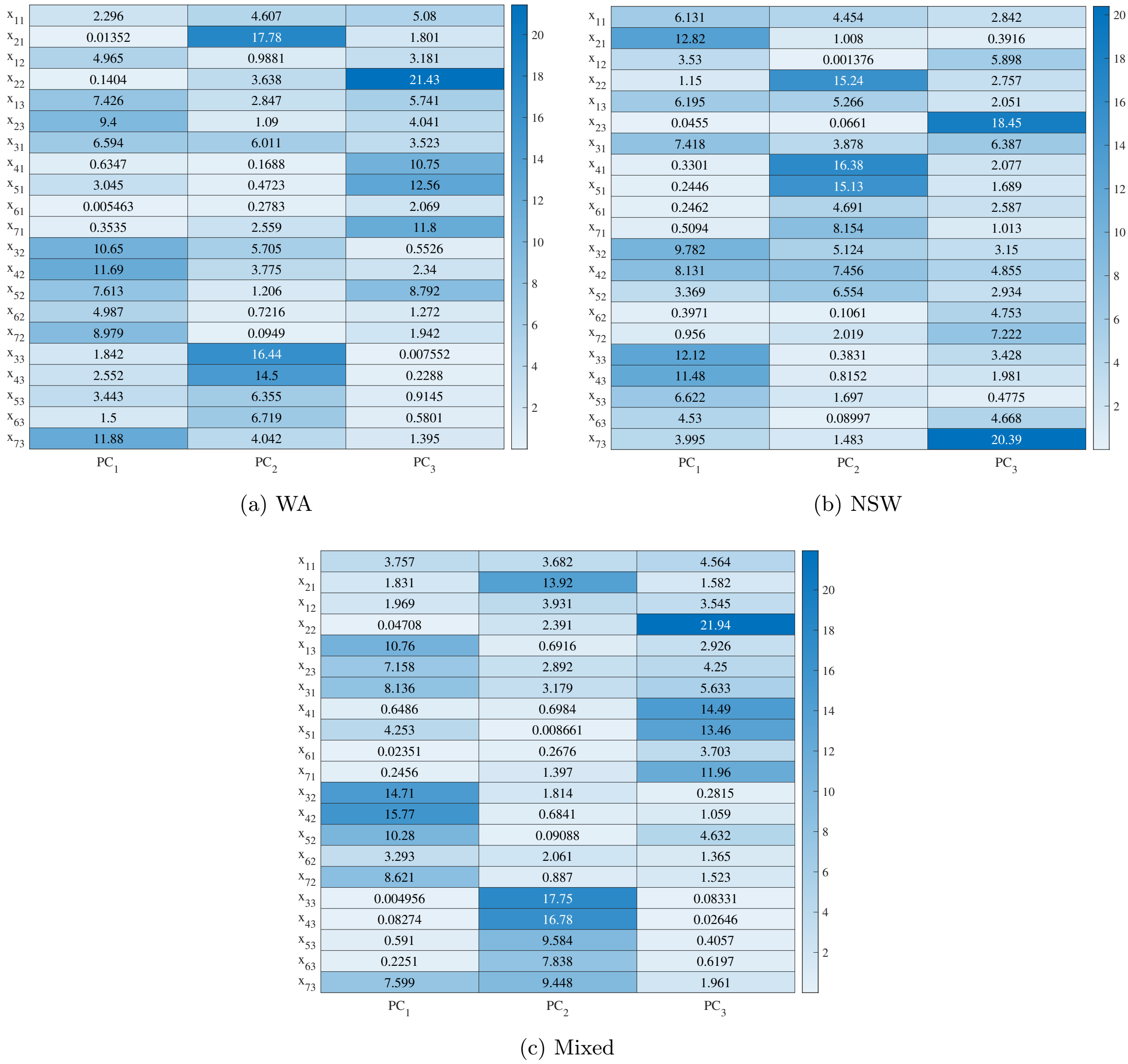


Figure 2: Loading matrix of features obtained from ultrasound tests conducted on billets at sites (a) WA, (b) NSW, and (c) mixed observations.

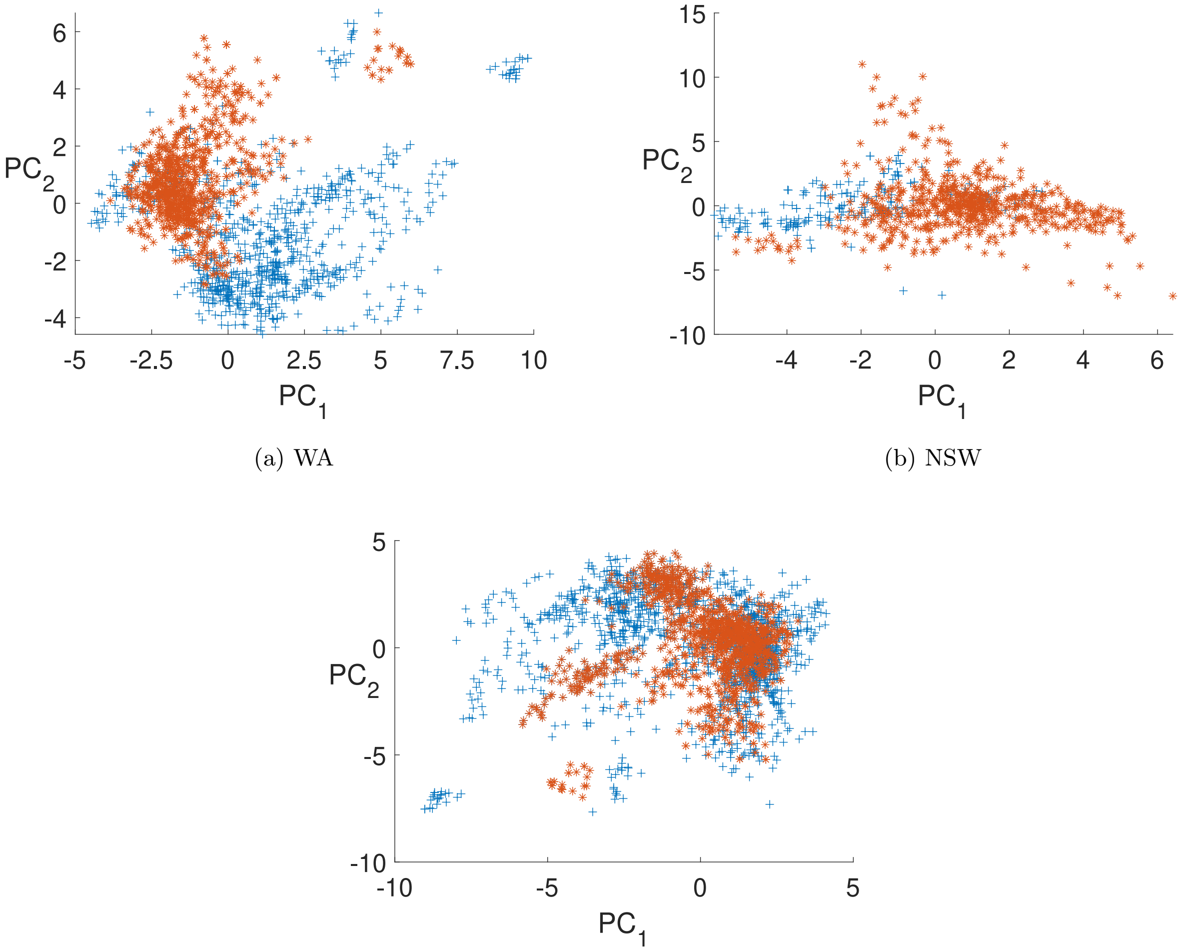
1. Calculate the proportion of variance accounted for by each PC obtained from the standardized feature matrix X.

2. Multiply the percentage of variance associated with each PC by the corresponding column of the matrix $\mathbf{Q}^2$.

3. Sum the results obtained in step (2) for the selected PCs. It is noteworthy that the selection of the number of PCs can be based on the cumulative variance explained by them. However, in this study, the first three PCs were consistently chosen.

Hence, the equation obtained for calculating the feature importance metric can be written as follows:

$$\mathbf{I} = \sum_{i=1}^{N} var(i) \times \mathbf{Q}^2(:, i) \tag{10}$$



(a) WA (b) NSW

(c) Mixed

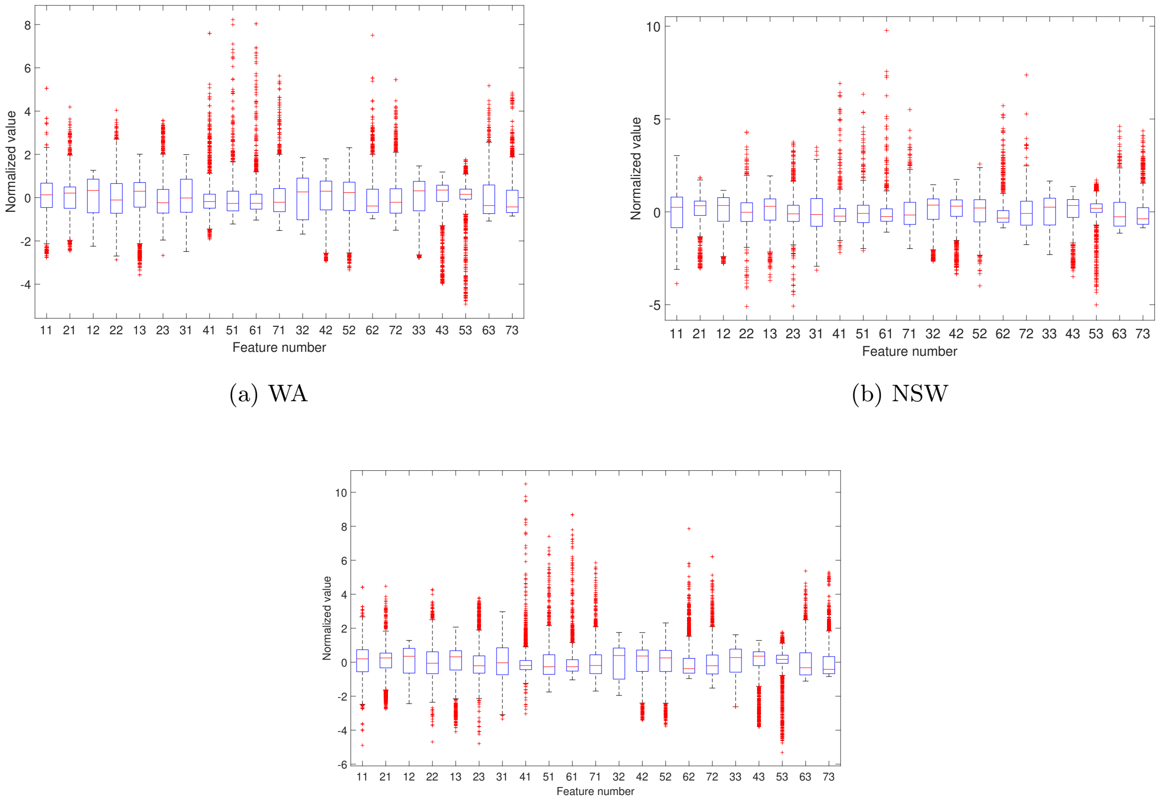
Figure 3: Scatter plot of PC_1 versus PC_2 of the standardized features obtained from ultrasound tests conducted at (a) WA, (b) NSW, and (c) mixed observations [1].

where $\mathbf{I}$ is a vector containing the importance values for features, N represents the number of Principal Components (PCs), the term $var(i)$ represents the proportion of variance accounted for by the i^{th} PC and N is set to 3.

4.1. The proposed MCD-based feature importance method

The proposed approach employs the Minimum Covariance Determinant (MCD) algorithm to robustly replace the factors $var(i)$ in (10). Specifically, the Fast Minimum Covariance Determinant (FastMCD) algorithm [14] is utilized to compute the robust covariance matrix for the feature matrix, ensuring a more resilient estimation of the variance factors. Such variance factors are the values located on the main diagonal of the robust covariance matrix. Basic details of the MCD algorithm are provided below for the reader's reference.

Consider the $n \times p$ feature matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ in which $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})$ represents the i^{th} observation. Notably, n and p indicate the number of observations and features. The aim is to identify the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, a vector of p components and a positive definite matrix



(a) WA

(b) NSW

(c) Mixed

Figure 4: The box plots depict the statistical distribution of standardized features obtained from ultrasound tests conducted on billets at three different sites: (a) WA, (b) NSW, and (c) mixed observations [1]. A number of outliers marked as "+" is evident from the plots.

of size $p \times p$, respectively, to represent the feature space $\boldsymbol{X}$ by an elliptically symmetric and unimodal distribution to obtain a subset of feature space that does not contain outliers.

As per the given definition, there must be a strictly decreasing real function g in the form of:

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{|\boldsymbol{\Sigma}|}}\, g\left(d^2\left(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)\right) \tag{11}$$

where $d(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ represnts the Mahalanobis distance calculated as

$$d\left(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})} \tag{12}$$

The MCD algorithm aims to obtain a subset of observations with the tuning parameter $n/2 \leq h \leq n$ to identify $(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0)$ so that [14],

1. $\hat{\boldsymbol{\mu}}_0$ is the average point of h optimal number of observations whose covariance matrix has the minimum determinant.

2. $\hat{\boldsymbol{\Sigma}}_0$ is the corresponding covariance matrix of the selected h observations multiplied by a consistency factor c_0.

3. The consistency factor c_0 is obtained as discussed in [15] and is meant to make the estimator consistent at the normal model.

4. The value of h needs to be chosen such that the determinant of the covariance matrix is not equal to zero. Specifically, h should be greater than $2p$. This condition implies that the number of observations, denoted by n, should be at least twice the number of dimensions, i.e., $n \geq 2p$. To minimize noise, it is recommended to have even more observations, specifically $n \geq 5p$, i.e., at least five incidents per dimension. The minimum regularized covariance determinant proposed in [16] can be used in higher dimensions.

5. The MCD estimator is known for its affine equivariance, meaning that it remains invariant under affine transformations. Put simply, for any non-singular matrix $\boldsymbol{A}p \times p$ and vector $\boldsymbol{b}p \times 1$, the MCD estimator maintains its properties.

$$\hat{\boldsymbol{\mu}}_{\text{MCD}}\left(\boldsymbol{X}\boldsymbol{A}^{'} + \mathbf{1}_n\boldsymbol{b}^{'}\right) = \hat{\boldsymbol{\mu}}_{\text{MCD}}\left(\boldsymbol{X}\right)\boldsymbol{A}^{'} + \boldsymbol{b} \tag{13}$$

$$\hat{\boldsymbol{\Sigma}}_{\text{MCD}}\left(\boldsymbol{X}\boldsymbol{A}^{'} + \mathbf{1}_n\boldsymbol{b}^{'}\right) = \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\text{MCD}}\left(\boldsymbol{X}\right)\boldsymbol{A}^{'} \tag{14}$$

   where $\mathbf{1}_n$ and $\boldsymbol{X}$ represents a column vector of n ones.

Hence, the equation for the proposed robust feature selection algorithm is a revised form of (10), which can be expressed as:

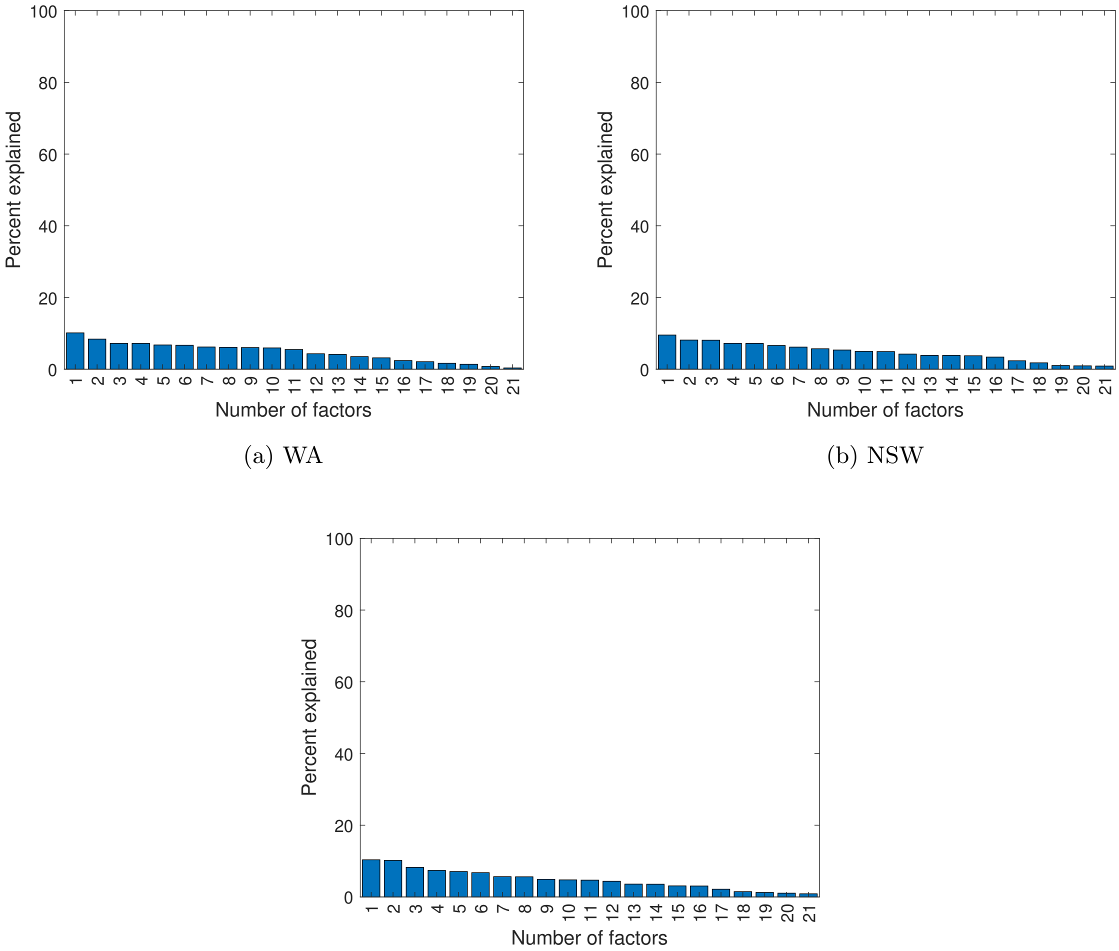
$$\mathbf{I} = \sum_{i=1}^{N} \boldsymbol{\Sigma}_{ii} \times \mathbf{Q}^2(:, i) \tag{15}$$

Here, $\boldsymbol{\Sigma}_{ii}$ denotes the elements located on the main diagonal of the robust covariance matrix. The equation calculates the sum of the element-wise product of $\boldsymbol{\Sigma}_{ii}$ and the i-th column of the loading matrix $\mathbf{Q}^2$. This modified equation serves as the foundation for the proposed robust feature selection algorithm.

5. Results and discussions

Figure 5 displays the variance percentage associated with each feature, as represented in the main diagonal of the robust covariance matrix obtained from the MCD algorithm applied to the feature matrices of specimens from WA, NSW, and mixed samples. The three highest values were subsequently utilized in (10) to calculate the importance of the features.

Figure 6 presents the top ten most important features obtained from the PCA and MCD-based techniques, ranked in descending order. As observed in the figure, the top ten selected features for WA specimens remain unchanged. However, there is a difference of one feature in the selection for NSW and mixed specimens. In the subsequent analysis, the ten most salient features extracted from the PCA and MCD-based methods were employed to address the classification challenges using diverse MLAs available in the machine learning toolbox of MATLAB. Table 5 presents a comparison of the performance of two different feature selection algorithms, namely PCA- and MCD-based techniques, in the context of field test results using the ten most important features. The results are reported for various MLAs applied



(a) WA

(b) NSW

(c) Mixed

Figure 5: The diagonal elements of the covariance matrix obtained from applying the MCD to the dataset for the following observations: (a) WA, (b) NSW, and (c) mixed observations.

to three different types of data: WA, NSW, and mixed data, obtained from the MCD- and PCA-based techniques.

The results show that in most cases, the performance of MLAs using features selected by MCD-based technique outperforms those selected by PCA-based technique. Specifically, for the NSW data, the classification results of MLAs using MCD-based features are consistently higher compared to those using PCA-based features, with some MLAs achieving statistically significant improvement in accuracy. The highest improvement is observed for the Fine Trees classifier, with an accuracy of 95.4% using MCD-based features compared to 94.1% using PCA-based features. Similar trends are observed for mixed data, where MLAs using MCD-based features generally outperform those using PCA-based features, albeit with varying degrees of improvement.

These results suggest that the MCD-based feature selection technique is more effective in identifying important features for classification compared to the PCA-based technique in the context of the field test results. The higher performance of MLAs using MCD-based features can be attributed to the robustness of MCD algorithm in identifying outliers and capturing non-linear relationships in the data, which may

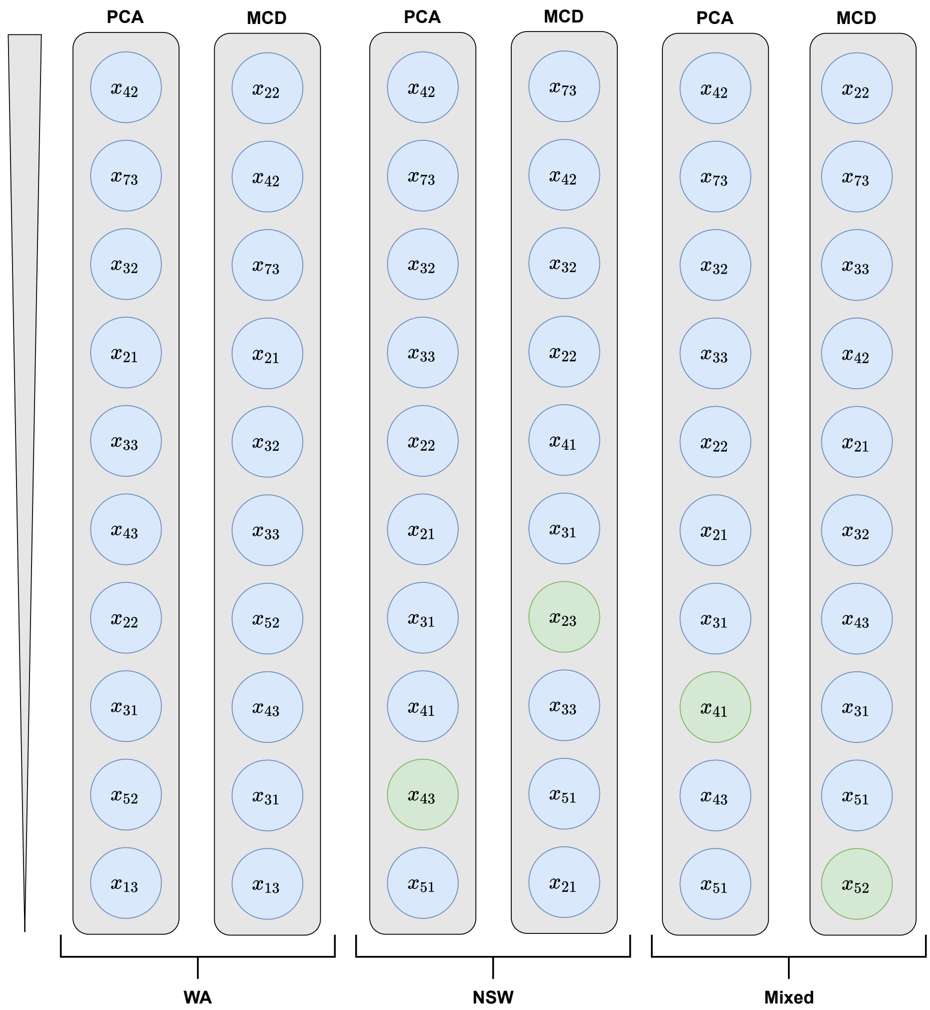


Figure 6: The top ten important features obtained from PCA- and MCD-based techniques.

not be adequately captured by PCA. These findings highlight the importance of careful selection of feature selection techniques in machine learning applications, and the need to consider the characteristics of the data and the specific problem at hand when choosing the appropriate feature selection algorithm. Further analysis and validation on larger datasets may be warranted to confirm these findings and generalize the results to other domains.

6. Conclusions

In conclusion, a robust feature importance ranking algorithm is proposed in this study to effectively handle outliers and capture non-linear relationships within a feature matrix. The proposed method utilizes the MCD algorithm to obtain the necessary factors for multiplying the loading matrix values in the feature importance equation. The results demonstrate that the top selected features differ when compared to an alternative PCA-based algorithm in cases involving NSW and mixed specimens. However, the top ten most important features for the WA data remained consistent. Notably, if a reduced number of features were used for training the MLAs, different features would be selected for the WA data.

Furthermore, the results of the five-fold classification using the MLAs indicate that the performance of the MLAs generally improved when incorporating the features selected by the MCD-based method

Table 5: The classification results using a 5-fold cross-validation for various MLAs applied to the field test results, utilizing the ten most significant features obtained from PCA- and MCD-based techniques.

MLA	**WA**	**NSW**	**Mixed**
	MCD/PCA	MCD/PCA	MCD/PCA
Fine trees	91.2 /91.2	**95.4**/94.1	**90.6**/90.3
Medium trees	91.0/91.0	**95.2**/94.5	**90.5**/89.8
Coarse Trees	88.2/88.2	**90.7**/90.6	**85.3**/85.2
Linear Discriminant	88.0/88.0	**90.6**/90.1	**86.6**/85.7
Quadratic Discriminant	87.5/87.5	**92.3**/91.3	**86.5**/85.7
Gaussian Naïve Byes	85.4/85.4	**85.3**/84.5	**83.3**/82.2
Kernel Naïve Byes	87.5/87.5	**88.1**/86.7	**86.3**/85.3
Linear SVM	90.4/90.4	**92.4**/91.9	**87.5**/86.8
Quadratic SVM	93.2 /93.2	95.1/**95.4**	**92.2**/91.6
Cubic SVM	91.8/91.8	**95.7**/95.4	**92.7**/92.2
Fine Gaussian SVM	93.9/93.9	**97.0**/96.7	**94.8**/94
Medium Gaussian SVM	93.3/93.3	95.2/95.2	**92.2**/91.7
Coarse Gaussian SVM	90.4/90.4	**90.5**/90.4	**88.1**/87.1
Fine Nearest Neighbor	91.2/91.2	**95.7**/94.8	**92.3**/91.5
Medium Nearest Neighbor	93.5/93.5	**95.8**/95.3	**93.1**/92.9
Coarse Nearest Neighbor	88.0/88.0	87.2/**87.5**	85.5/**85.7**
Cosine Nearest Neighbor	92.2/92.2	**94.8**/93.6	**92.2**/91.8
Cubic Nearest Neighbor	93.5/93.5	**95.1**/94.9	**92.9**/92.8
Weighted Nearest Neighbor	93.3/93.3	95.6/**95.9**	**94.4**/94.1
Boosted Trees	92.7/92.7	**95.8**/94.9	**93.1**/92
Bagged Trees	93.3/93.3	96.1/**96.4**	**93.9**/93.3
Subspace Discriminant	86.6/86.6	90.1/90.1	**85.7**/85.2
Subspace KNN	91.5/91.5	**96.9**/96.5	**93.3**/92.2
RUSBoosted Trees	91.9/91.9	**95.9**/95.5	**91.3**/90.7

into the models. This highlights the significance of employing a robust feature importance algorithm when training machine learning models with data that exhibit a high number of outliers and non-linear dependencies among features.

It is important to note that the proposed feature importance method is unsupervised, as it does not rely on labels corresponding to the features. While the findings of this study are promising, it is necessary to conduct further comparisons between the proposed method and similar alternatives to validate the results and extend their applicability to other domains. Furthermore, the proposed technique still relies on the assumption of data following a Gaussian distribution. Therefore, future work needs to be focused on modifying the proposed method to make it applicable to diverse datasets.

References

1. Mohsen Mousavi, Mohammad Sadegh Taskhiri, and Amir H Gandomi. Standing tree health assessment using contact–ultrasonic testing and machine learning. Computers and Electronics in Agriculture, 209:107816, 2023.

2. Mark A Hall. Correlation-based feature selection for machine learning. PhD thesis, The University of Waikato, 1999.

3. Kenji Kira and Larry A Rendell. A practical approach to feature selection. In Machine learning proceedings 1992, pages 249–256. Elsevier, 1992.

4. Carsten F Dormann, Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R García Marquéz, Bernd Gruber, Bruno Lafourcade, Pedro J Leitão, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36(1):27–46, 2013.

5. Utkarsh Mahadeo Khaire and R Dhanalakshmi. Stability of feature selection algorithm: A review. Journal of King Saud University-Computer and Information Sciences, 34(4):1060–1073, 2022.

6. Erick Odhiambo Omuya, George Onyango Okeyo, and Michael Waema Kimwele. Feature selection for classification using principal component analysis and information gain. Expert Systems with Applications, 174:114765, 2021.

7. Mohsen Mousavi, Mohammad Sadegh Taskhiri, Damien Holloway, JC Olivier, and Paul Turner. Feature extraction of wood-hole defects using empirical mode decomposition of ultrasonic signals. NDT & E International, 114:102282, 2020.

8. Mohsen Mousavi and Amir H Gandomi. Wood hole-damage detection and classification via contact ultrasonic testing. Construction and Building Materials, 307:124999, 2021.

9. Mohsen Mousavi, Amir H Gandomi, Damien Holloway, Adam Berry, and Fang Chen. Machine learning analysis of features extracted from time–frequency domain of ultrasonic testing results for wood material assessment. Construction and Building Materials, 342:127761, 2022.

10. Dennis Gabor. Theory of communication. part 1: The analysis of information. Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering, 93(26):429–441, 1946.

11. Ivanovich Muskhelishvili, Nikolaĭ and Jens Rainer Maria Radok. Singular integral equations: boundary problems of function theory and their application to mathematical physics. Courier Corporation, 2008.

12. Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. IEEE Transactions on Signal Processing, 62(3):531–544, 2014.

13. Dominique Zosso. Variational mode decomposition, matlab central file exchange, Retrieved August 27, 2020. URL https://www.mathworks.com/matlabcentral/fileexchange/44765-variational-mode-decomposition.

14. Mia Hubert, Michiel Debruyne, and Peter J Rousseeuw. Minimum covariance determinant and extensions. Wiley Interdisciplinary Reviews: Computational Statistics, 10(3):e1421, 2018.

15. Christophe Croux and Gentiane Haesbroeck. Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. Journal of Multivariate Analysis, 71(2):161–190, 1999.

16. Kris Boudt, Peter J Rousseeuw, Steven Vanduffel, and Tim Verdonck. The minimum regularized covariance determinant estimator. Statistics and Computing, 30(1):113–128, 2020.