ORIGINAL ARTICLE
Journal Section

Prediction of wind turbine generator failure using 2 stage cluster-classification methodology

Alan Turnbull[1†] | Dr James Carroll[1] | Dr Alasdair McDonald[1] | Sofia Koukoura[1]

[1]Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1XW, UK

Correspondence
Alan Turnbull, Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1RD, UK
Email: a.turnbull@strath.ac.uk

Present address
[†]Electronic and Electrical Engineering, Technology and Innovation Centre, 99 George Street, Glasgow, G1 1RD, UK

Funding information
EPSRC, project reference number EP/L016680/1

Reducing wind turbine downtime through innovations surrounding asset management has the potential to greatly influence the LCoE for large wind farm developments. Focusing on generator bearing failure and vibration data, this paper presents a two-stage methodology to predict failure within 1 to 2 months of occurrence. Results are obtained by building up a database of failures and training machine learning algorithms to classify the bearing as healthy or unhealthy. This is achieved by first using clustering techniques to produce sub-populations of data based on operating conditions, which this paper demonstrates can greatly influence the ability to diagnose a fault. Secondly, this work classifies individual clusters as healthy or unhealthy from vibration based condition monitoring systems by applying order analysis techniques to extract features. Using the methodology explained in the report, an accuracy of up to 81.6% correct failure prediction was achieved.

KEYWORDS
Wind turbine, failure, bearing, condition monitoring, vibration, machine learning

Abbreviations: LCoE; Levelised cost of energy, O&M; operation and maintenance, CM; condition monitoring, SCADA;supervisory control and data acquisition, SVM; support vector machine, WT; wind turbine.

1 | INTRODUCTION

In order for wind energy to compete with traditional methods of generating electricity such as fossil fuels, the levelised cost of energy (LCoE) must be further reduced in the coming years. Costs associated with the operation and maintenance (O&M) of a wind farm makes up a significant proportion of total lifetime costs. In fact, up to 30% of the total energy cost can be spent on O&M for some large offshore developments [1]. With wind farms moving into harsher environments further offshore, this value is only expected to increase in the future. As more money is spent on O&M, innovations surrounding asset management have the potential to greatly influence the overall LCoE. According to a study found in [2], innovations associated with operations, maintenance and service are anticipated to reduce the LCoE by approximately 2% between 2014 and 2025. Generator faults can contribute significantly to the overall downtime experienced by a wind farm due to component failure, with around 1 failure per year in state of the art offshore wind turbines (WT) [3] [4] [5] [6] [7] [8].

One of the areas in which significant improvement can be made is through the introduction of turbine condition-based maintenance [2]. All large utility scale WTs have supervisory control and data acquisition (SCADA) systems as standard, which are primarily used for performance monitoring [9], however developers are now opting for more sophisticated condition monitoring (CM) systems to gain better insight into WT condition. The most widely developed and adopted CM systems are based on vibration analysis, with sensors placed throughout the nacelle to gain insight into the dynamic performance of WT sub-systems, and in turn identify any potential issues or faults. This paper will focus on WT generator bearing failure of a multi-megawatt machine, and there are a number of well established and proven techniques using vibration analysis to detect faults in such systems. These are traditionally based around Fourier analysis, Hilbert transform and order analysis techniques, which can all be used to identify and extract features from the signal based on rotational frequencies [10].

With increased wind farm capacity and asset portfolios growing in size there is a growing need for automated fault detection, which can utilise proven techniques and flag issues in real time across an entire fleet. The research presented in this paper draws upon synchronised databases of generator bearing vibration time series and failure events from a wind turbine original equipment manufacturer (OEM). This allows multiple vibration signal examples of the same failure mode in different turbines at a number of time intervals leading up to failure to be analysed and compared. This paper will first of all introduce the bearing failure selected in this study and briefly describe the techniques used to detect faults based on spectral analysis, providing insight into the key features which can be identified and extracted. Secondly, this paper will briefly introduce the principles behind using classification algorithms to predict failure based on extracting features from vibration measurements before presenting prediction results.

Previous studies have shown that operating parameters at the time of vibration measurement can have a substantial effect on vibration spectra, and hence the features which are extracted to detect the fault [11] [12]. To help mitigate these issues vibration samples can be binned based on power output at the time of measurement, a technique widely adopted by developers of condition monitoring software today. Expanding on this idea of grouping vibration data, this paper presents a two-stage methodology to predict generator bearing failure 1-2 months before occurrence. The first stage will use *k*-means clustering to group data by operating conditions, which will act as an advancement to existing power binning techniques by considering more parameters to define measurement sample bins. The second stage uses decision trees and support vector machines for feature classification. *K*-means clustering separates data based on Euclidean distance from a set number of cluster centroids, therefore does not discriminate clusters based on sample size. Both Decision Tree's and SVM's are well established techniques, with the former scaling well to larger groups and the latter known to perform better with smaller groups [13] [14].

This paper provides insight into how clustering techniques can be used to improve classification accuracy by

FIGURE 1 Damaged shaft from inner ring of bearing spinning on it (left) and damage to inner ring from spinning on shaft (right).

separating vibration samples based on key operating conditions. The study uses a significantly larger pool of real world data and examples of the same generator bearing failure in identical machines than any other previous study to date, allowing for a comprehensive analysis of how failures observed across a fleet can be used to detect the same failure in other similar WTs. It also presents additional evidence to support the hypothesis that operating conditions have a substantial effect on the ability to detect faults and predict failure, and presents a robust and scale-able method to separate vibration measurements prior to classification, with the highest accuracy achieved at mid-range operating conditions for an example wind turbine and failure mode.

2 | GENERATOR BEARING FAILURE

This paper will focus on generator bearing failure, which in this case stems from raised bearing temperatures leading to bearing inner ring growth resulting in the bearing inner ring spinning on the generator rotor shaft at the drive end. For this reason, the fault leading up to failure will be treated as a mechanical looseness issue within the internal assembly, with Figure 1 showing damage that can occur over time from a fault of this nature. Root causes sometimes stem from design and manufacturing issues such as imperfections in material grade, out of tolerances and improper installation methods. Other causes include operational and maintenance issues such as high loading, unbalanced electromagnetic forces, damage while in transit or exacerbated by inadequate cooling and inspection strategies [15].

3 | METHODOLOGY

Figure 3 shows the overall methodology used in this study, with the initial stage compromising of the data gathering and labelling process. Next the vibration signals were analysed and the key features were extracted which could indicate the fault or describe the operating conditions at the time of measurement. Two models are presented in this paper, both of which used classification algorithms to classify the bearing as healthy or unhealthy. The first acting as a baseline model in which to compare the second two stage cluster-classification model.

3.1 | Data population

In order to track and compare component condition, vibration data was retrieved at different points in time leading up to a generator bearing failure. To achieve this, events associated with generator bearing failure from a wind turbine OEM were analysed until several examples of the same failure mode were identified. This was then cross checked with

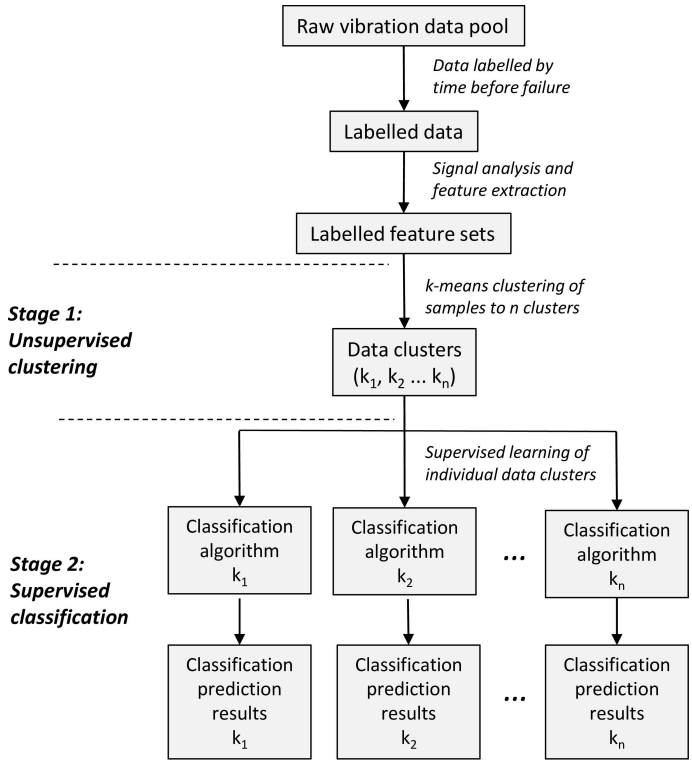

FIGURE 2 Two stage methodology for training and validating algorithms to predict generator bearing failure using *n* number of clusters.

SCADA data to ensure dates were correlated by checking failure dates with down time experienced by the wind turbine. To guarantee a fair comparison all examples were from an identical generator and drivetrain configuration. In order to ensure confidentiality the exact power output and wind turbine and bearing type used is not provided, however it can be stated that it was a doubly-fed induction generator (DFIG) with a rated power of between 2 and 4 MW. Each turbine utilised a variable speed, pitch regulated control strategy. Generator rotor speed at rated power was partly determined by grid frequency, where examples were found for both 50 and 60 Hz.

A total of 15 different wind turbines from eight wind farms were used in the study, in each of which between 4 and 10 vibration samples were gathered a week apart 1 year, 5-6 months and 1-2 months before failure. In most cases data was gathered from each of the three groups described above. In some cases however not every group could be provided, in which case data was used where available. This is clarified in Table 1. Each sample consisted of approximately 10 seconds of data taken with a sampling frequency of approximately 25 kHz at both the drive end and non-drive end generator bearing (see Figure 3 for clarification).

Once the population of failure examples had been identified the data gathered prior to occurrence, data was labeled based on the time to failure. A binary classification system was used, with samples labeled 'healthy' taken at more than 2 months prior to failure and labeled 'unhealthy' if the sample was taken between failure and 2 months prior to failure. This gave a total of 306 vibration samples, consisting of 204 in class A (healthy) and 102 in class B (unhealthy) for both the drive end and non-drive end sensor.

TABLE 1 Pool of data available for analysis

Wind turbine No.	Healthy - at least 1 year	Healthy - 5-6 months	Unhealthy - 1-2 months
1	10 samples 1 week apart	9 samples 1 week apart	7 samples 1 week apart
2	None	None	4 samples 1 week apart
3	9 samples 1 week apart	7 samples 1 week apart	8 samples 1 week apart
4	8 samples 1 week apart	7 samples 1 week apart	8 samples 1 week apart
5	10 samples 1 week apart	10 samples 1 week apart	7 samples 1 week apart
6	9 samples 1 week apart	8 samples 1 week apart	8 samples 1 week apart
7	8 samples 1 week apart	7 samples 1 week apart	7 samples 1 week apart
8	8 samples 1 week apart	8 samples 1 week apart	8 samples 1 week apart
9	9 samples 1 week apart	9 samples 1 week apart	6 samples 1 week apart
10	None	9 samples 1 week apart	7 samples 1 week apart
11	None	9 samples 1 week apart	7 samples 1 week apart
12	9 samples 1 week apart	9 samples 1 week apart	5 samples 1 week apart
13	8 samples 1 week apart	8 samples 1 week apart	6 samples 1 week apart
14	None	None	7 samples 1 week apart
15	7 samples 1 week apart	9 samples 1 week apart	7 samples 1 week apart
Total	**95 samples**	**109 samples**	**102 samples**

3.2 | Feature extraction

Most of the experience to draw upon to analyse vibration in generators and other rotating machinery comes from industries which utilise large, fixed speed synchronous machines. Modern wind turbines employ variable speed control strategies and, along with the stochastic nature of the wind, produce load patterns that are far more varied than traditional generators. The analysis of such vibration signals are therefore more challenging and as such, makes diagnosing faults in wind turbine generators more difficult.

3.2.1 | Time domain analysis

The vibration signal can be analysed in a number of ways, the simplest of which is in the time domain. Basic statistical analysis techniques can provide important information about the signal and although it is not sufficient to actually diagnose faults on its own, it certainly provides a useful method in which to detect any obvious irregularities. Table 2 shows the features that were extracted and used to analyse the signal, where $x(t)$ is the acceleration at time t, x_{max} and x_{min} are the maximum and minimum measured acceleration, u_x, x_{rms}, σ_x and β are the mean acceleration, RMS, standard deviation and kurtosis over signal length T.

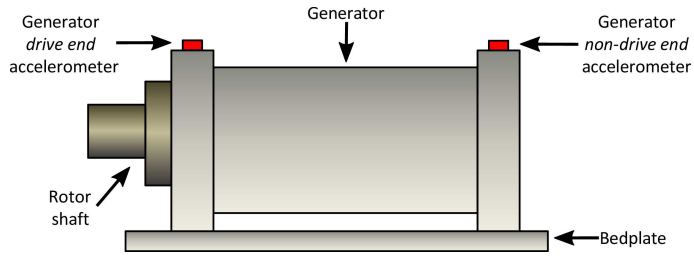

FIGURE 3 Diagram showing estimated positions of accelerometers used to measure generator bearing vibration.

TABLE 2 Time-domain features

Feature No.	Feature	Formula
1.1	Maximum	$x_{max} = max\{x(t)\}$
1.2	Minimum	$x_{min} = min\{x(t)\}$
1.3	Mean	$u_x = \frac{1}{T}\int_0^T x(t)dt$
1.4	RMS	$x_{rms} = \left[\frac{1}{T}\int_0^T x^2(t)dt\right]^{1/2}$
1.5	Standard deviation	$\sigma_x = \sqrt{\frac{1}{T}\int_0^T [x(t)-u_x]^2 dt}$
1.6	Kurtosis	$\beta = \frac{1}{T}\int_0^T [x(t)-u_x]^4 dt$

3.2.2 | Fourier & order analysis

The fastest and most widely used technique for analysing vibration signals is through the Fourier Transform, which represents the signal in the frequency domain, in which spectral peaks and signal energy can be analysed [16]. However, during each vibration sample, the generator shaft speed can vary, often significantly, meaning that the signal is not stationary. This can produce a smearing effect on the FFT spectrum which is somewhat proportional to the range of shaft speeds experienced over the sample. The Fourier transform can therefore be adapted for a sliding time window by using the short-time Fourier transform, where a spectro-temporal representation of the signal is obtained. This is used for order analysis and allows the spectral values to be tracked in time [17] [18]. Order analysis is a re-sampling technique which can be effective when analysing non-stationary signals. When a signal is re-sampled using this technique it moves from a variable shaft speed and constant sampling rate, to constant shaft speed and variable sample rate. When computing the frequency-RPM map it is important that a sensible resolution bandwidth is chosen to capture all the desired features of the signal therefore losing as little information as possible. Table 3 shows the features that were extracted and used to analyse the signal using this technique. The spectral peak frequencies were chosen to reflect amplitude gains expected for a bearing fault such as this causing relative movement within the internal assembly of the generator [12] [15] [19] [20].

Although all WTs in the dataset were the same size and topology, they did not all use the same bearing manufacturer and model. This meant that the ball passing frequency for the inner and outer races (BPFI and BPFO), as well as the rolling element deterioration frequency (BSF) could not be calculated consistently for each WT. For this reason only frequencies which were common to all WTs were used in the analysis, which included synchronous vibration at rotor speed fundamental frequency (Order No. 1) along with the 2nd (Order No. 2) and 3rd (Order No. 3) harmonics. Non-synchronous vibration indicators were also extracted at order numbers 0.5, 1.5 , 2.5 and 3.5.

TABLE 3 Order-domain features

Feature No.	Feature	Description
2.1	$A_{0.5}$	Amplitude of peak at order number 0.5
2.2	A_1	Amplitude of peak at order number 1
2.3	$A_{1.5}$	Amplitude of peak at order number 1.5
2.4	A_2	Amplitude of peak at order number 2
2.5	$A_{2.5}$	Amplitude of peak at order number 2.5
2.6	A_3	Amplitude of peak at order number 3
2.7	$A_{3.5}$	Amplitude of peak at order number 3.5

TABLE 4 Operational characteristics

Feature No.	Feature	Description
3.1	P_{out}	Average electrical power out during 10s signal
3.2	Ω_{gen}	Average shaft speed during 10s signal
3.3	τ_{gen}	Average electromagnetic torque during 10s signal $\tau_{gen} = P_{out}/\Omega_{gen}$
3.4	U_{wind}	Average wind speed during 10s signal from wind turbine anemometer

3.2.3 | Operational characteristics

Operational characteristics are used as a method of understanding the loading conditions at the time of vibration measurement. The features that were extracted to represent this are shown in Table 4 and were used in two instances; first of all in the baseline model to assist with training the classification algorithms, and secondly in two-stage model used independently from classification during the clustering stage. A wind turbine control strategy is defined on a torque-speed curve while the operating performance is indicated by the power output in relation to wind speed. By considering these 4 features a representation of both relationships can be achieved.

3.3 | Classification models and prediction algorithms

3.3.1 | Baseline classification model

The baseline classification model trained the classification algorithms with all available data regardless of turbine and operating conditions at the time of vibration measurement. Both support vector machine (SVM) and decision tree classifiers were tested, with decision trees proving to be best suited to the baseline model due to their ability to scale easily to large data populations. Once algorithms were trained they could be validated using 5-fold cross validation (see Section 3.4). This provided a baseline to which the two-stage methodology could be fairly compared to evaluate the effectiveness of clustering data by operating conditions to produce sub-populations of data.

3.3.2 | Two-stage classification model

For the two-stage model, as shown in Figure 2, features were clustered using *k*-means clustering to create groups containing vibration samples taken at similar operating conditions. The same algorithms as described above were then trained on features within each cluster, with the best algorithm chosen based on performance, which changed with the number and spread of samples in that particular group. Once algorithms were trained they could be validated using 5-fold cross validation, before the chosen algorithm was exported for each cluster. Once exported the chosen algorithm could be used for any unseen data belonging to any particular cluster. Testing SVM's and decision trees provided the opportunity for high levels of accuracy across clusters containing both small and large quantities of data, which will be discussed in more detail in Section 4.

3.3.3 | Prediction algorithms

The classification algorithms described in this Section are used across both the baseline and two-stage model, while clustering is only required in the latter. All algorithms are widely known techniques and can be readily applied.

There are many types of common clustering algorithms which can find hidden patterns in large data sets, each of which use different metrics to analyse the data. Such algorithms include Hierarchical clustering, *k*-Means clustering, Hidden Markov models and Gaussian mixture models, however for this paper *k*-means clustering was chosen due to its ease of application, fast computational speed and ability to choose the number of clusters [21]. *K*-means clustering works by partitioning data into distinct clusters based on the euclidean distance to the centroid of each cluster. Any number of clusters can be used, which is predefined (not chosen by the algorithm like many other techniques) prior to partitioning.

Once features were successfully grouped, classification algorithms were then trained and tested based on the validation process outlined in Section 3.4, with the best chosen and applied to the set of features specific to that cluster. There are a variety of classifiers available that use supervised learning processes to classify data. For this research decision trees and support vector machines with polynomial kernel were chosen. Decision trees scale well making them suitable for datasets with a large number of samples, while SVM's typically are more suitable for datasets with a smaller number of samples. In such cases however, they provide an opportunity for more complex decision boundaries to be established [21].

Decision trees are simple algorithms which classify data based on a variety of features. At the root of the tree all features are considered before the data is split into branches based on predetermined conditions. This process of using logic to split data continues using a series of internal nodes until branches can no longer be split again, producing a leaf, or final decision at the terminal node. The process of splitting data is key to producing a useful decision tree, as each split will have a cost associated in terms of performance and accuracy. Too many splits can cause trees to be overly complex and results in data overfitting, meaning it does not generalise the data well. Decision trees can be refined by pruning methods, which remove branches utilising features of low importance, and boosting methods, which increase tree stability, or the likelihood of small changes making a significant impact on the overall tree design. SVM's are adaptable algorithms widely used for both classification and regression problems. They work by plotting each data point in *n*-dimensional space, with *n* being the number of features used to train the model. Classification is achieved by finding the hyper-plane that differentiates the classes of coordinates (also known as support vectors). The margin is a measure of the distance between the hyper-plane and nearest support vector of each classification. A large margin indicates that the support vector machine is stable and will be less susceptible to misclassifying data [22]. Both classifiers are well established and more information can be found about their application in [13] [23].

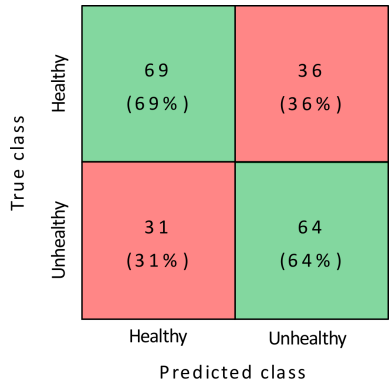

FIGURE 4 Confusion matrix of trained algorithms for all data (100 healthy and 100 unhealthy samples).

3.4 | Validation of results

There are two key stages when applying machine learning to vibration data for classification; a training phase followed by a validation phase. This process was repeated with different permutations of features to discover which could best be used to produce the greatest overall accuracy. For each labeled vibration sample leading up to failure the features described in Tables 2 to 4 were extracted and used. These were then clustered with data in each cluster according to Table 5. To ensure the algorithm is trained in a balanced manner, the same number of data points were chosen from each class, with random samples chosen for the class with more samples in any particular group. For example, if one cluster contained 40 healthy samples and 32 unhealthy samples, 32 samples for each health class were used to train the model. Cross-validation was used to determine the overall accuracy of the algorithm. This method involves partitioning the data into subsets of a predetermined ratio, one of which is then omitted from training and used to test the algorithm. For this example 20% of the data was used for validation purposes. The process is then repeated using different sup-populations and an average accuracy calculated to use as a performance indicator. The prediction process can then be evaluated using a confusion matrix giving correct/incorrect classification and the likelihood of false positives/ negatives.

4 | RESULTS & DISCUSSION

4.1 | Baseline classification model

The baseline model had a maximum classification accuracy of 67% using a decision tree classifier. The confusion matrix is presented in Figure 4, which shows the likelihood of false positives or negatives when classifying the bearings as healthy or unhealthy. The data population used for both training and validation was limited only by the number of available unhealthy examples. The results shown here used 100 random vibration samples from each health class and regardless of the random population of data used, the accuracy never increased beyond 67%, with the mean accuracy just over 64% for this particular classification algorithm. The fact that accuracy changes substantially based on the data chosen for training highlights the limitations of this approach and suggests that a more robust methodology is required to choose data more carefully for supervised training.

The results above used all order domain features described in Table 3, along with the time series RMS, standard deviation and kurtosis (features 1.4-1.6) in Table 2. For the baseline model, operational characteristics outlined in Table 4 were also utilised for classification.

TABLE 5 Cluster descriptions

Cluster No.	Data in cluster (%)	Centroid position ($\tau_{gen}, U_{wind}, U_{gen}$)	Cluster description
1	15.0	(0.093, 0.19, 0.61)	low torque, low wind speed, low gen speed
2	4.2	(0.23, 0.25, 0.72)	low-med torque, low-med wind speed, low-med gen speed
3	20.5	(0.27, 0.31, 0.87)	medium torque, medium wind speed, medium gen speed
4	41.0	(0.38, 0.38, 0.99)	medium torque, rated wind speed, high gen speed
5	19.3	(0.83, 0.63, 0.99)	rated torque, rated wind speed, high gen speed

4.2 | Two-stage classification model

4.2.1 | Clustering results

K-means clustering partitions data using an iterative process initiated with a random guess of centroid positions. This means that although the number of clusters can remain constant, different data clusters are obtained every time the algorithm is ran with a different seed. A sensitivity study performed using 10 random seeds with 5 clusters showed that 96% of the data points were consistently grouped into a particular cluster. Results will focus on 5 clusters, chosen to provide a balance between separating the vibration samples based on operating conditions while still having enough data in each cluster to perform classification and validation. Using 5 clusters also manages to cover all parts of the torque-speed operating curve including important transition points. The effect of number of clusters on classification accuracy is presented in more detail in Section 4.2.3.

Samples were clustered based on three properties which describe the operating conditions at the time the signal was taken; average electromagnetic torque (feature 3.3), wind speed (feature 3.4) and generator speed (feature 3.2). The average shaft speed and torque gives a good indication of loading conditions on the generator shaft at the time of measurement, while using wind speed allows for an understanding of any de-rating or adjustments from the normal torque speed curve operating points. Only 3 out of the 4 parameters described in Table 4 were required due to the direct relationship between power and torque. For this analysis using electromagnetic torque provided more consistency in the clustering process. The clusters are described qualitatively in Table 5, while Figure 5 shows the data clustered into the 5 groups to assist with visualisation. Each point in Figure 5 represents the coordinates of the chosen operation characteristics used for clustering for a single vibration sample. Note that these characteristics are common to both the drive-end and non drive-end sample. The centroid position is also given in Table 5, which describes each cluster quantitatively with respect to the normalised wind speed, electromagnetic torque and generator speed.

4.2.2 | Classification results

The results presented for the two-stage classification model use the 5 clusters of data shown in Section 4.2.1. Each cluster goes through a training and validation process for each of the classification algorithms as described in Section 3.3.3. Table 6 presents the best algorithm and accuracy for each of the clusters, along with the percentage accuracy

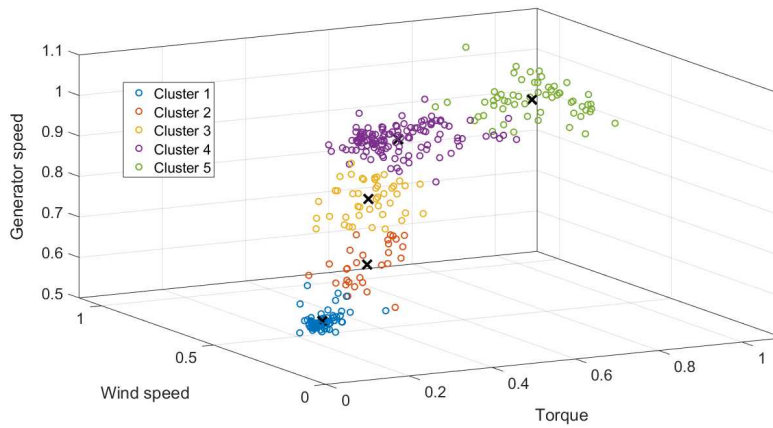

FIGURE 5 Plot of clusters based on three chosen variables, normalised to ensure confidentiality. 'X' represents each cluster centroid position.

TABLE 6 Results for 5 clusters

Cluster No.	Algorithm	Overall accuracy (%)	Difference from base model (%)
1	Decision Tree	61	-6
2	SVM	70	+3
3	Decision Tree	81.6	+14.6
4	Decision Tree	78.6	+11.6
5	Decision Tree	65.2	-1.8

difference from the baseline model described in Section 4.1, for which all data was taken and classified without any clustering.

Two of the groups showed a significant increase in accuracy with a maximum of up to 81.6% for group 3. These two groups (3 and 4) accounted for 61.5% of all data, while group 2, which saw a comparatively modest rise of 3% in accuracy accounted for a further 4.2%. Interestingly, the two groups (1 and 5) which underperformed in comparison to the baseline model had the 2 extreme datasets of either very high operating conditions or very low. This suggests that if the operating conditions are too low at the time of vibration sample it becomes more difficult to distinguish between a healthy or unhealthy bearing. Alternatively, if the WT is operating at rated power with rated torque, shaft speed and wind speed, this also holds true. The algorithms perform much better using features from mid-range operating conditions, with the highest accuracy when in and around rated wind speed. Figure 6 shows the overall accuracy of the algorithms for each group in comparison to the baseline, with the groups ordered based on the extremity of operating conditions at the time the vibrations sample was taken.

In terms of features analysis, the dominant frequency (the point at which the most increase in amplitude was observed to indicate the fault) changed between turbines, therefore all order domain features were used in every case. The RMS, standard deviation and kurtosis proved to be the most significant of the time domain features, with accuracy generally increasing if at least one of these three features was used.

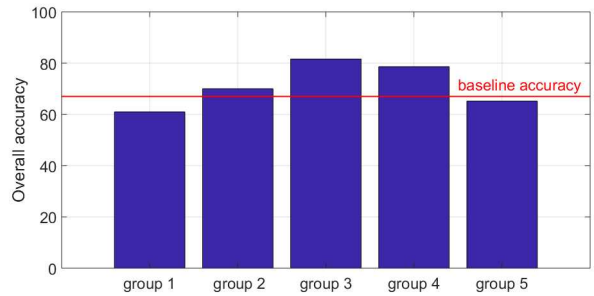

FIGURE 6 Plot of accuracy for each cluster ordered by extremity of operating conditions at time of vibration sample.

To compare absolute accuracy of both approaches, an overall weighted accuracy for the two-stage model can be calculated by;

$$\gamma_{overall} = \sum_{i}^{n} \gamma_i \delta_i \tag{1}$$

where $\gamma_{overall}$ is the overall weighted accuracy, γ_i is the cluster accuracy and δ_i is the percentage of data in cluster i for n number of clusters. For the 5 clusters in this analysis this gives an overall weighted accuracy of 73% for the two-stage methodology considering all data. If groups 1 and 5 are excluded, the weighted accuracy of the remaining groups increases to 79%, and represents all data in mid-range operating conditions.

Looking at the best performing algorithms in Table 6, decision trees consistently outperformed SVM's in the 4 groups with the highest number of data samples. As expected, SVM's performed better in group 2, which had significantly less data. Having the option of both algorithms proved to be a useful method to ensure the highest accuracy was achieved for each group of varying size.

Considering now the top two performing groups in more detail, Figure 7 shows the confusion matrix for each training and validation phase. Results indicate a balance between predicting false-positives and false-negatives, with each group having slightly less misclassification's for unhealthy samples to that of healthy. Fault propagation times will differ for every wind turbine used in the study, therefore the features extracted as fault indicators will also change in time leading up to failure. Classifying all data 1-2 months before failure as unhealthy is one of the limitations surrounding this type of classification process, which considers multiple examples without making an allowance for the fact each bearing has its own unique propagation time and failure threshold.

4.2.3 | Effect of number of clusters on classification accuracy

The results highlighted in Section 4.2.2 showed the two-stage classification model with 5 clusters. As stated previously, this provided a balance between the number of clusters and number of data within each cluster. If 3 or 4 were chosen an increase in accuracy at mid range operating conditions still occurred, however the same level of improvements were not observed. Tables 7 and 8 show the breakdown of clusters and accuracy along with the associated cluster centroid positions and algorithms used. Considering a weighted accuracy as described in Section 4.2.2, this can be calculated

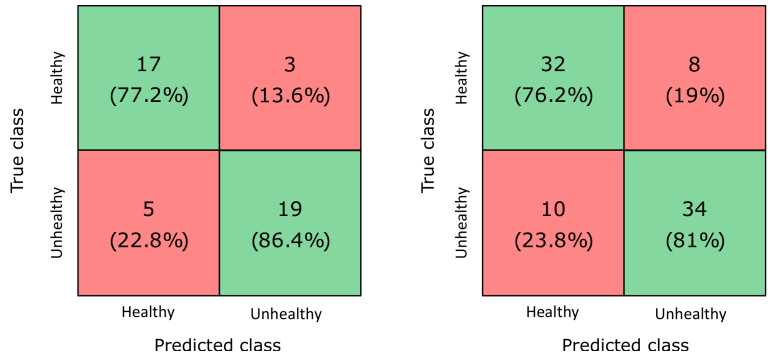

FIGURE 7 Confusion matrix of trained algorithms for group 3 (left) and group 4 (right).

TABLE 7 Results for 3 clusters

Cluster No.	Data (%)	Centroid	Algorithm	Overall accuracy (%)
1	31.1	(0.13, 0.21, 0.64)	Decision Tree	62
2	50.1	(0.34, 0.36, 0.96)	Decision Tree	78.2
3	18.8	(0.81, 0.62, 0.99)	SVM	60.4

TABLE 8 Results for 4 clusters

Cluster No.	Data (%)	Centroid	Algorithm	Overall accuracy (%)
1	30.3	(0.13, 0.20, 0.63)	Decision Tree	62.5
2	43.6	(0.31, 0.35, 0.95)	Decision Tree	79.7
3	14.5	(0.60, 0.48, 0.99)	SVM	63.0
4	11.6	(0.90, 0.68, 0.99)	SVM	53.6

for 3 clusters as 69.8% and 69.1% for 4 clusters. If 6 or more clusters were chosen, groups started to get too sparse to perform classification on every cluster, (as even with 5 clusters group 2 consisted of only 4.2% of the data) and accuracy generally decreased. If more data was available, it would be worth revisiting the number of clusters to determine the optimal range.

4.2.4 | Comparison of clustering and power binning

Up to this point clustering has been compared only to the baseline model, which considers all available data. Clustering offers the opportunity to group vibration data using a variety of operating parameters, and results have shown improved accuracy in doing so, however, it is also important to compare this approach to standard power binning practices. A similar methodology was used as previously describe in Figure 2, but instead of clustering in stage 1 data was divided into 5 even bins based on power output only. Each bin was then classified individually, with results showing an improvement to the baseline model, but not reaching the level of accuracy achieved through clustering. Results are presented in Table 9, where mid-range operating conditions again provide the highest prediction accuracy.

TABLE 9 Results by power binning

Bin No.	Power range (%)	Data in bin (%)	Algorithm	Overall accuracy (%)
1	0-20	18.3	Decision Tree	63.6
2	20-40	45.8	Decision Tree	68.2
3	40-60	15.8	Decision Tree	71.1
4	60-80	5.9	SVM	67.1
5	80-100	14.2	SVM	55.2

5 | CONCLUSION

Predictive maintenance strategies that use previous failures to learn and predict failure and remaining useful life of components in different wind turbines have the potential to make substantial savings to costs associated with O&M. This research indicates that machine learning classification algorithms can be applied to specific features to successfully predict generator bearing failure 1-2 months before occurrence with an accuracy of up 81.6%. To achieve this level of accuracy operating conditions at the time each vibration sample was taken must be considered independently of classification, as accuracy falls significantly to 67% without it, as shown in the baseline model. A simple approach to improve baseline accuracy is manual or rule-based removal of data at extremes of the power curve, however it is difficult to do this robustly and fairly. Binning vibration samples by power output to create groups which can be fairly compared is standard practice, however this paper has demonstrated that *k*-means clustering is a more successful method by considering more operating parameters prior to classification.

This work shows that the operating conditions at the time of vibration measurement can greatly influence the ability to detect a fault, with mid-range operating conditions leading to highest accuracy when predicting remaining useful life based on vibration analysis for this particular wind turbine and failure mode. The examples shown in Section 4 find a balance between false-positives and false-negatives, meaning that equal weight is given to either case however from a commercial perspective it may be worth refining the model to either minimise or maximise false-positives, which could be further explored through cost-benefit analysis.

The two-stage methodology described in this paper provides a scalable prediction model which could be applied to a range of faults across the drivetrain. By considering different features of vibration that can effectively describe a particular fault, which may also rely on different vibration measurements and analysis techniques to those used in this paper, the model could be adapted accordingly.

REFERENCES

[1] Crabtree CJ, Zappala D, Hogg SI. Wind energy: UK experiences and offshore operational challenges. Proceedings of the Institution of Mechanical Engineers Part a-Journal of Power and Energy 2015;229(7):727–746.

[2] Valpy B, English P. Future renewable energy costs: offshore wind, How technology innovation is anticipated to reduce the cost of energy from European offshore wind farms; 2014.

[3] Carroll J, McDonald A, McMillan D. Reliability Comparison of Wind Turbines With DFIG and PMG Drive Trains. Ieee Transactions on Energy Conversion 2015;30(2):663–670.

[4] Hahn B, Durstewitz M, Rohrig K. Reliability of wind turbines - Experiences of 15 years with 1,500 WTs. Wind Energy 2007;p. 329–+.

[5] Pfaffel S, Faulstich S, Rohrig K, Performance and Reliability of Wind Turbines: A Review. MPDI energies; 2017.

[6] Spinato F, Tavner PJ, van Bussel GJW, Koutoulakos E. Reliability of wind turbine subassemblies. Iet Renewable Power Generation 2009;3(4):387–401.

[7] Carroll J, McDonald A, McMillan D. Failure rate, repair time and unscheduled O and M cost analysis of offshore wind turbines. Wind Energy 2016;19(6):1107–1119.

[8] Wilkinson M, Harman K, Spinato F, Hendriks B, Delft TV, Measuring Wind Turbine Reliability - Results of the Reliawind Project; 2011.

[9] Tautz-Weinert J, Watson S. Using SCADA data for wind turbine condition monitoring – a review. IET Renewable Power Generation 2017;11(4):382–394.

[10] Mollasalehi E. Data-driven and Model-based Bearing Fault Analysis - Wind Turbine Application. Thesis; 2017.

[11] de Azevedo HDM, Araujo AM, Bouchonneau N. A review of wind turbine bearing condition monitoring: State of the art and challenges. Renewable Sustainable Energy Reviews 2016;56:368–379.

[12] Hossain ML, Abu-Siada A, Muyeen SM. Methods for Advanced Wind Turbine Condition Monitoring and Early Diagnosis: A Literature Review. Energies 2018;11(5):14. `<GotoISI>://WOS:000435610300287`.

[13] Stetco A, Dinmohammadi F, Zhao XY, Robu V, Flynn D, Barnes M, et al. Machine learning methods for wind turbine condition monitoring: A review. Renewable Energy 2019;133:620–635.

[14] Saidi L, Ali JB, Fnaiech F. Application of higher order spectral features and support vector machines for bearing faults classification. ISA Transactions 2015;54:193 – 206. `http://www.sciencedirect.com/science/article/pii/S0019057814002043`.

[15] Girdhar P, Scheffer C, Practical Machinery Vibration Analysis and Predictive Maintenance; 2004.

[16] Chen JL, Pan J, Li ZP, Zi YY, Chen XF. Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals. Renewable Energy 2016;89:80–92.

[17] Kabir MJ, Oo AMT, Rabbani M, Ieee. A Brief Review on Offshore Wind Turbine Fault Detection and Recent Development in Condition Monitoring Based Maintenance System. In: Australasian Universities Power Engineering Conference (AUPEC) Australasian Universities Power Engineering Conference, NEW YORK: Ieee; 2015. `<GotoISI>://WOS:000376680000085`.

[18] Romero A, Soua S, Gan TH, Wang B. Condition monitoring of a wind turbine drive train based on its power dependant vibrations. Renewable Energy 2018;123:817–827.

[19] Borghesani P, Pennacchi P, Chatterton S, Ricci R. The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery. Mechanical Systems and Signal Processing 2014;44(1):118 – 133. Special Issue on Instantaneous Angular Speed (IAS) Processing and Angular Applications.

[20] Wang Y, Tse PW, Tang BP, Qin Y, Deng L, Huang T, et al. Order spectrogram visualization for rolling bearing fault detection under speed variation conditions. Mechanical Systems and Signal Processing 2019;122:580–596. `<GotoISI>://WOS:000457948600029`.

[21] Muller AC, Guido S, Introduction to Machine Learning with Python. O'Reilly Media; 2017.

[22] Laouti N, Sheibat-Othman N, Othman S. Support vector machines for fault detection in wind turbines. IFAC Proceedings Volumes (IFAC-PapersOnline) 2011;18(PART 1):7067–7072.

[23] Kateris D, Moshou D, Pantazi XE, Gravalos I, Sawalhi N, Loutridis S. A machine learning approach for the condition monitoring of rotating machinery. Journal of Mechanical Science and Technology 2014;28(1):61–71. `https://doi.org/10.1007/s12206-013-1102-y`.

ALAN TURNBULL received the M.Eng degree in Aero-mechanical Engineering from the University of Strathclyde in 2013. He is a Ph.D. student at the University of Strathclyde, focusing on developing diagnostic and prognostic methods for wind turbine generators by applying signal processing and machine learning techniques.

DR JAMES CARROLL received the B.Eng. degree in mechanical and manufacturing engineering from Dublin Institute of Technology in 2008, and the M.Eng. in Technical Management from Emden University of Applied Science, Germany in 2010. He received a Ph.D. from the University of Strathclyde in 2016. He is a lecturer in the EPSRC Wind Energy Doctoral Training Centre, University of Strathclyde.

DR ALASDAIR MCDONALD received the M.Eng degree in Electrical and Mechanical Engineering from the University of Durham, U.K., in 2004 and a Ph.D. degree from the University of Edinburgh, U.K. in 2008. He is a lecturer in the EPSRC Wind Energy Doctoral Training Centre, University of Strathclyde.

SOFIA KOUKOURA received her degree in Mechanical Engineering from the National Technical University of Athens in 2015. She then joined the University of Strathclyde Wind and Marine Energy Systems Centre for Doctoral Training. Her PhD focuses on diagnostics and prognostics of wind turbine gearboxes using signal processing and machine learning techniques.