applied sciences

MDPI

Article

Remaining Useful Life Estimation of Rotating Machines through Supervised Learning with Non-Linear Approaches

Eoghan T. Chelmiah *,†, Violeta I. McLoone and Darren F. Kavanagh *,†

Faculty of Engineering, Institute of Technology Carlow, R93 V960 Carlow, Ireland; violeta.mcloone@itcarlow.ie
* Correspondence: eoghan.chelmiah@ieee.org (E.T.C.); dfkavanagh@ieee.org (D.F.K.)
† These authors contributed equally to this work.

Abstract: Bearings are one of the most common causes of failure for rotating electric machines. Intelligent condition-based monitoring (CbM) can be used to predict rolling element bearing fault modes using non-invasive and inexpensive sensing. Strategically placed accelerometers can acquire bearing vibration signals, which contain salient prognostic information regarding the state of health. Machine learning (ML) algorithms are currently being investigated to accurately predict the health of machines and equipment in real time. This is highly advantageous towards reducing unscheduled maintenance, increasing the operational lifetime, as well as mitigation of the associated health risks caused by catastrophic machine failure. Motivated by this, a robust CbM system is presented for rotating machines that is suitable for various industrial applications. Novel non-linear methods for both feature engineering (one-third octave bands) and wear-state modelling (exponential) are investigated. The paper compares two main types of feature extraction, which are derived from Short-Time Fourier Transform (STFT) and Envelope Analysis (EA). In addition, two types of supervised learning, Support Vector Machines (SVM) and *k*-Nearest Neighbour (*k*-NN) are explored. The work is tested and validated on the PRONOSTIA platform dataset, with remaining useful life (RUL) classification results of up to 74.3% and a mean absolute error of 0.08 achieved.

Keywords: mechanical bearings; signal processing; fault detection; feature extraction; frequency domain analysis; machine learning; condition-based monitoring; rotating machine; Support Vector Machine (SVM); *k*-Nearest Neighbour (*k*-NN)

check for updates

Citation: Chelmiah, E.T.; McLoone, V.I.; Kavanagh, D.F. Remaining Useful Life Estimation of Rotating Machines through Supervised Learning with Non-Linear Approaches. *Appl. Sci.* **2022**, *12*, 4136. https://doi.org/10.3390/app12094136

Academic Editors: David Mba and Tat-Hean Gan

Received: 12 March 2022
Accepted: 12 April 2022
Published: 20 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Electric machines are a vital component for major industries, such as manufacturing, mining, agriculture, energy and transport sectors. It is fair to say that all of these sectors are currently undergoing major growth and technological innovation. These machines are typically required to operate under harsh environmental conditions and demanding drive cycles, which gives rise to premature degradation and the occurrence of catastrophic failure modes [1,2]. It is imperative for the future viability and sustainability of these sectors that we have efficient, robust and highly reliable electric machines. Sudden catastrophic machine breakdown results in acute manufacturing downtime, dramatic reductions in productivity and health and safety concerns. Moreover, performing critical maintenance is both labour-intensive and costly. Faults can be difficult to diagnose and troubleshoot for maintenance teams [3–5].

Broadly, electric machines can be broken down into the bearings, stator, rotor and other elements as shown in Figure 1a. The statistics of failure for three classes, low voltage, medium voltage and high voltage, are presented in Figure 1b. Bearings are the dominant failure mode at low and medium voltages, followed closely in the latter class by stator fault modes. The final class, high voltage, is dominated by stator fault modes, as seen in Figure 1b. Research studies have shown bearings to be responsible for up to 75% of low-voltage electric machines breakdowns and up to 41% of all rotating machine failures [2,6–9].

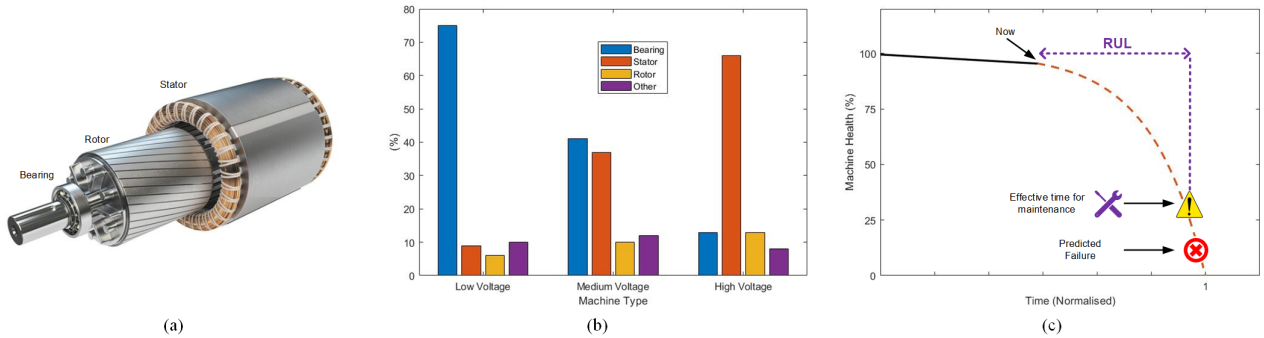


Figure 1. (**a**) The main components of an electric machine, (**b**) percentage occurrences of failure modes from the statistical data presented in [2,6] and (**c**) typical degradation trend and remaining useful life of an operating machine.

Condition-based monitoring has received considerable attention over the past years [2,10–12]; hence, a rich literature exists at present. The areas of detection and diagnosis of fault modes has received the most attention with mature industrial technologies available. More recently, advanced methods of prognosis are being investigated, and these focus on the more challenging problem of predicting the Remaining Useful Life (RUL) of the machines or sub-components [13–15]. This is illustrated in Figure 1c. Knowing the RUL of a component ensures that the state of health of a machine is known and that suitable maintenance can be performed at the optimal times. The maximum usable life is achieved without the threat of total machine breakdown occurring [16–19].

Typically, machine-learning (ML) methods for CbM focus on three key aspects: signal sensing modalities, feature engineering and supervised learning algorithms. Sensing modalities for CbM methods that have been explored in recent years include acoustic emission (noise) [5,20–22], air gap torque monitoring [23,24], motor current signal analysis (MCSA) [25,26], oil and debris-monitoring [27,28], thermography [29,30], speed fluctuations monitoring [31], induced voltage, magnetic flux monitoring and vibration signals [4,5,11,16].

The Short-Time Fourier Transform (STFT) feature extraction method has been extensively used to extract useful time-frequency features, which reported to achieve high levels of classification accuracy [32–35]. Envelope Analysis (EA) has also been used extensively for prognostic and diagnostic purposes as the method is simple yet versatile, making it applicable to many different types of mechanical fault monitoring processes [21,22,36–38]. Other methods of time-frequency feature extraction, which have shown good promise for bearing prognostic use cases are Wavelet Transformation (WT) [39–42] and Empirical Mode Decomposition (EMD) [43–48], both of which were reported to achieve highly accurate performance scores.

In the area of supervised learning, two of the most versatile, simple and effective ML techniques that have been extensively used for bearing condition monitoring are *k*-Nearest Neighbour (*k*-NN) [49–51] and Support Vector Machines (SVM) [52–55]. More recently, advanced deep-learning techniques have been applied for prognostic purposes, and these use multiple traditional methods combined and neural networks to perform RUL classification [40,45,48,56–59].

In [59], an RUL prediction method was proposed based on a long short-term memory (LSTM) neural network framework and deep features, which were learned adaptively from the two health states. Benali [48] proposed a method to characterize and classify seven different bearing classes using statistical features, EMD energy entropy and an artificial neural network (ANN). Li [56] used a combination of two supervised ML techniques; a regression model and multi-layer ANN to predict the RUL of rolling element bearings.

Here, in this paper, a novel ML method for RUL is proposed, using non-linear signal processing techniques to perform feature engineering based on STFT and EA with One-third octave band feature compression. The rationale for using Fourier and EA-based

feature extraction was as a result of detailed vibration signal analysis conducted on the bearing signals from the dataset. This motivated the incorporation of non-linear feature compression of the multidimensional feature space using Octave bands as the prognostic information signatures are highly concentrated in the lower portion of the spectra.

These features form part of the ML method recipes alongside twelve different supervised learning algorithms based on *k*-NN and SVM to determine the optimal choice for RUL classification. SVM and *k*-NN algorithms were chosen because of their robustness for supervised learning problems, in particular problems with datasets of limited size, which greatly limits the suitability of applying other more advanced deep-learning approaches, e.g., ANN and LSTM. The work also highlights the importance of using non-linear wear-state models to track the degradation severity levels; this has been shown to greatly improve the performance of the ML classifiers overall for this RUL task. The time frequency analysis conducted on the vibration signals also motivated the investigation of non-linear wear state models as the bearing degradation typically does not follow linear trends.

The remainder of this paper is organised as follows. Section 2 presents a graphical and statistical analysis of the vibration signals. The proposed ML method is detailed in Section 3. The experimental procedure is illustrated in Section 4. In Section 5, the results from the proposed CbM system are presented. Section 6 presents the major trends and findings from the results and statistical analysis, and the limitations of this work are discussed. Finally, Section 7 concludes the work and highlights future research avenues to explore.

2. Vibration Signal Analysis

The typical degradation of bearings is a gradual and slowly evolving process. Ordinarily, it would take many years to acquire signals characterising the entire process, starting with a new bearing and progressing to a fully degraded bearing at the end of its life cycle. Typically, applied experiments are performed in the laboratory under controlled conditions to acquire bearing data. These accelerated ageing experiments can involve artificially inducing faults by strategically drilling small holes or etching the surface of the bearings, applying excessive loads and operating speeds, or elevating the temperature and humidity, as reported in [60,61].

2.1. Dataset

The proposed condition monitoring methods described in this research were tested and validated using the award winning FEMTO-ST Institute 2012 PHM Challenge dataset [62], which has found widespread use in this field [63–65]. This complete dataset comprises of vibration signals from the accelerated degradation of 17 bearings, performed at three different operating speed and load conditions: (1800 rpm and 4000 N), (1650 rpm and 4200 N) and (1500 rpm and 5000 N) [62]. This dataset provides realistic fault modes achieved under accelerated ageing conditions as opposed to data generated from artificial fault modes, e.g., drilled holes, machined narrow cuts or indentation lines to emulate the occurrence of hairline cracks.

Accordingly, for each test case, the details on the specific type of failure mode or element (e.g., ball, inner race or outer race or cage) is not known or provided. This work is more concerned with monitoring degradation rather than diagnosis of the specific failure mode. The test setup was composed of three main parts: a rotating part, a degradation generation (loading) part and a sensing measurement part. The bearing's vibration amplitudes were recorded by two miniature accelerometer sensors, Dytran, Model 3035B, placed orthogonally to one another on the vertical and horizontal axis of the bearing under test. This sensor pair was placed radially on the external race of the bearing. The acceleration measures were sampled by the analogue to digital converter (ADC) at 25.6 kHz [62].

Each recording consists of bursts of 2560 samples (0.1 s duration), which were obtained every 10 s throughout the test bearings' lifetimes up until the point of failure. Failure for these test bearings was defined to be the point where the amplitude of the vibration signals surpassed the reference acceleration threshold value of 20 g. This threshold was carefully

chosen as it also avoided any considerable propagation that could severely damage the test-bed mounts and fixtures [62].

2.2. Signal Analysis

The testing and validation of the proposed wear state estimation system was focused towards the first operating condition, with speeds of 1800 rpm and an applied force of 4000 N, for seven bearing test cases. The duration of each test case varied in length with the longest test case reaching almost 8 h and the shortest lasting only 2 h and 25 min as illustrated in Table 1 and Figure 2. Since the run-to-failure test was conducted with no artificial mechanical tampering of the constituent bearings, it can be expected that a spread of different fault types would have occurred in the bearings, involving the rolling elements, the cage and the inner-race and outer-race parts.

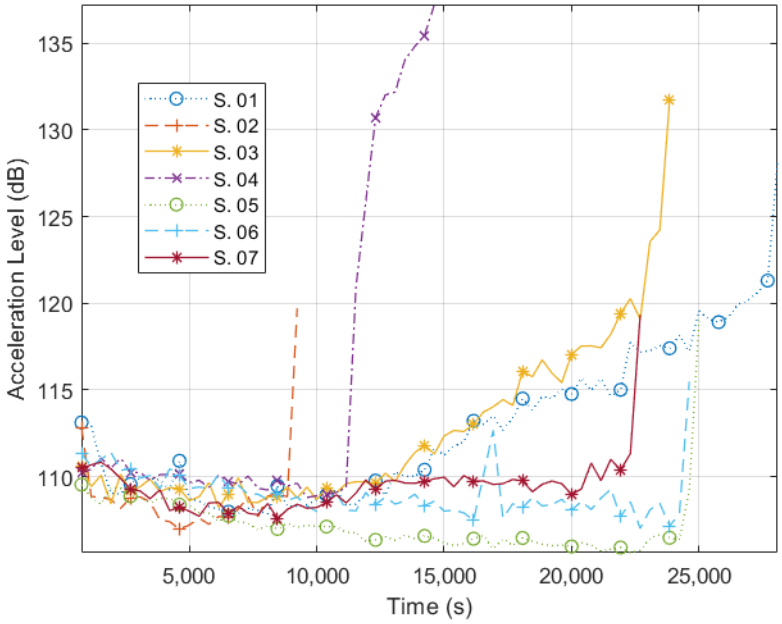


Figure 2. Moving average (MA) plots of 7 vibration signals are shown in the form of acceleration level (dB) across time (s), for condition 1 of the bearing dataset [62]. The MA interval is 512 points and a reference level of 1 $\mu m/s^2$ was used, as per [66].

Table 1. Timing characteristics of datasets.

Sample Bearing (S.)	Lifetime Duration (Hour | Min | Sec)
S. 01	7 h 47 m 00 s
S. 02	2 h 25 m 00 s
S. 03	5 h 00 m 10 s
S. 04	3 h 09 m 40 s
S. 05	6 h 23 m 30 s
S. 06	6 h 23 m 29 s
S. 07	4 h 10 m 11 s

Figure 3 demonstrates examples of typical time domain vibration signals where different fault types occur. The signal depicted in the three top panels show a very gradual increase in the vibration amplitude for bearing S. 01 before the fault occurs. The fault in bearing S. 04, depicted in the three lower panels, manifests itself as a sudden change in the vibration amplitude about three quarters through the lifetime. This indicates the occurrence of a very different, abrupt fault mode, such as the rapid formation of a catastrophic crack, a part snapping or sudden deformation due to heat induced by friction.

These two examples, that of the slowly evolved degradation mode and the rapid formation fault mode, go some way to demonstrate the inherent degree of complexity and

challenges that exist in developing robust condition-monitoring systems using machine-learning methods. It is difficult to ascertain trends and patterns in the time domain signals for the vibration amplitudes alone. Hence, a conversion to the frequency domain is necessary to observe spectral signatures and trends throughout the duration of the test to failure. Accordingly, a Short-Time Fourier Transform (STFT) process was applied to each bearing test case with the parameters set to produce a multivariate spectral description of the data.

Table 2 details the STFT algorithm parameters used in this work. The concept of STFT analysis is fundamental for describing any quasi-stationary (slowly time varying) signal. In general, one can define the STFT in terms of the output of an arbitrary bank of filters. The amplitude spectrum of each frequency component of the signal was converted to a decibel scale.

Figure 2 illustrates the variance of mean STFT spectral component energies for each time sample for all 7 of the bearing test cases, using the first 50 time samples of each as their reference baseline value. The natural degradation trend of a bearing does not represent a gradual, linear pattern as represented in Figure 2.

All seven bearings test cases vary in the number of time samples as no two experiments were equal in duration, however, the avalanche-like pattern of degradation is apparent in all cases. This sudden effect makes the RUL estimation more difficult as the wear states leading to this stage share a great deal of the same feature values, and some do not vary from their initial baseline values until roughly 75% of their usable lifetime has passed.

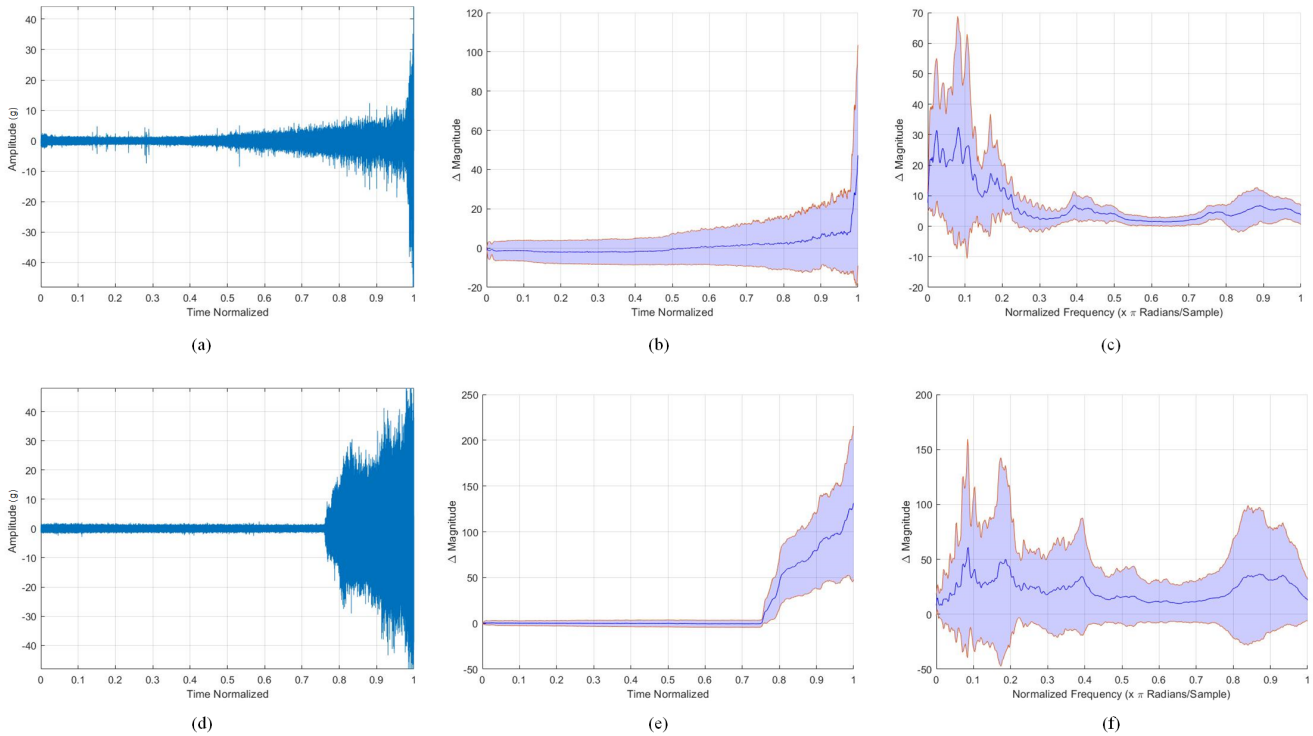


Figure 3. Plots (**a**,**d**) illustrate the time series vibration signals of bearings S. 01 and S. 04, respectively. Panes (**b**,**e**) show the variance of the total frequency amplitude values of each signal acquisition over time. The shaded error bar illustrates the mean ± one standard deviation. Panes (**c**,**f**) show the variance of the frequency variance of bearing S. 01 and S. 04 throughout the entire ageing process. The shaded error bar illustrates the mean frequency spectral amplitudes ± one standard deviation.

Table 2. STFT parameters.

Sampling Frequency (f_s)	25,600 Hz
Window Length (L_w)	1024
Hop Size (L_h)	256 (75%)
No. of FFT Points ($nfft$)	1024
Window Type (w)	Blackman
Amplitudes	dB

3. Proposed ML Method

This study investigated and compared a variety of algorithm options for bearing wear-state classification. The proposed method begins by taking raw accelerometer data and concludes with assigning a predicted wear state class. A number of intermediate steps are involved. This section describes each method stage and describes how each stage can branch into alternative steps as illustrated in Figure 4.

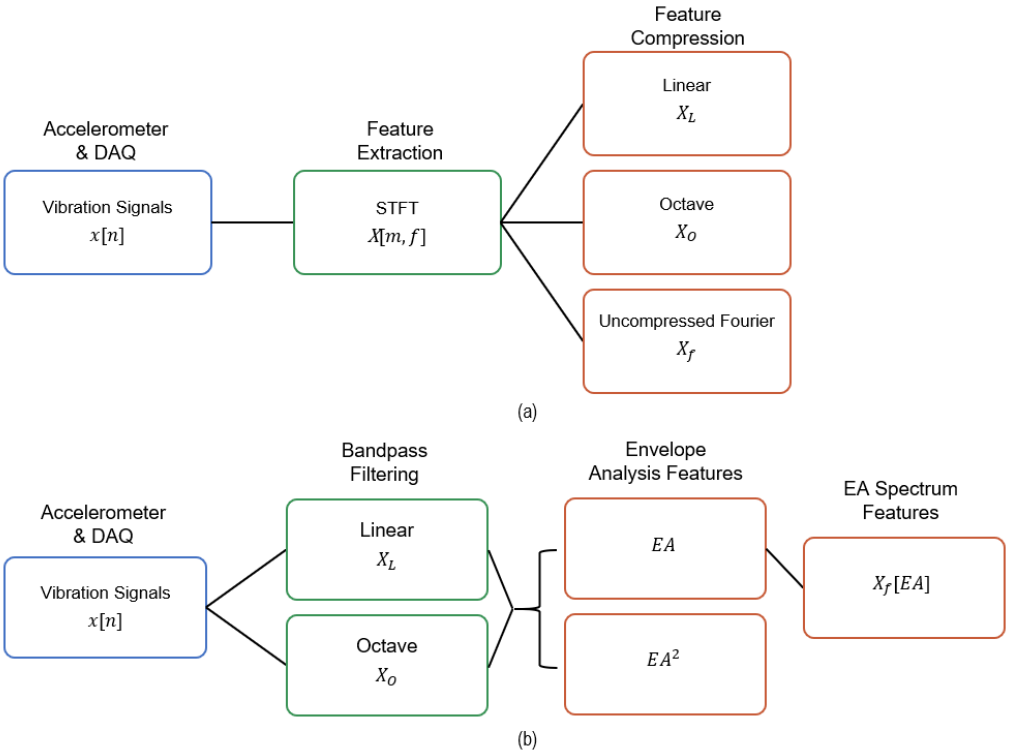


Figure 4. (**a**) STFT feature extraction approach with feature compression and (**b**) the three feature extraction approaches investigated for Envelope Analysis (EA).

3.1. Feature Extraction

The proposed method begins with a feature extraction step and two feature extraction methods were performed for comparison. The two techniques of extracting classification features from the raw time-series data were: (1) applying a STFT to the discrete-time signals and (2) calculating the signal envelopes of each discrete-time signal.

3.1.1. Discrete-Time Signal Analysis

The non-stationary time series data recorded from the 2012 PHM Data Challenge bearing dataset are presented as a 2-D vector. The vibration amplitude sampled at 25.6 kHz by the Dytran Model 3035B accelerometer was transferred from the discrete-time domain to the frequency domain using the Short-Time Fourier transform (STFT) parameters detailed in Table 2.

The Short-Time Fourier transform (STFT) can be defined as a sequence of Fourier transforms of a windowed input signal. STFT is used to extract time-localized frequency

information, for situations in which frequency components of a signal vary drastically over time, such as non-stationary bearing vibration signals [67]. The STFT, shown in Equation (1), involves calculating a windowed Fast Fourier transform (FFT) of the discrete time samples, with each window overlapping the previous by a factor of 75% to obtain the complex feature vector signatures across time. The average value of these absolute complex feature vectors for each sample is calculated to extract a spectral feature vector consisting of 512 spectral points (bins).

The spectral points are spaced at 25 Hz intervals and represent the spectral amplitudes content over a range of 0 to 12,800 Hz (Nyquist frequency). This frequency range is determined by specifying the number of Discrete Fourier transform (DFT) points calculated for each window and by the sampling frequency of the DFT. The sampling frequency was matched to the frequency of the accelerometers sampled the bearing signals at, 25.6 kHz, and the number of DFT points calculated for each window was 1024 points (the same value as the length of the sample window $g(n)$).

$$\mathcal{F}\Big\{x[n]\Big\} = X[m,f] = \sum_{n=-\infty}^{\infty} x[n]g[n-mR]e^{-j2\pi fn} \tag{1}$$

where $\mathcal{F}\Big\{x[n]\Big\}$ is the DFT of windowed data centred about time mR, $g(n)$ is a window function, and R is the hop size between successive DFTs.

The STFT has been favoured as a feature extraction method to obtain useful features for both RUL estimations and fault classification and achieved extremely high results as seen in [34,35,68,69].

3.1.2. Envelope Analysis

An Envelope Analysis (EA) approach, involving the extraction of the signal envelope of the discrete-time vibration signals, was also used as an alternative method to extract time-frequency features. Two different filtering approaches were applied to the non-stationary signals, linear and non-linear. The linear filtering process uses equidistant frequency spacing, whilst the non-linear option involves applying the one-third octave band scale to the vibration signal in the time domain. This produced 25 representations of the signal for both the linear and the non-linear cases for classification performance comparison. The filter employed was a 25 Finite Impulse Response (FIR) sixth order Butterworth filter.

The discrete-time domain signal is demodulated by taking the absolute value of the non-stationary signal points, to produce $x_r[n]$. By taking the Hilbert Transform of this rectified signal, we can produce the signal, $x_i[n]$, which enables the creation of the complex analytic signal, defined as $z(n)$ and shown in Equation (2).

$$z[n] = x_r[n] + jx_i[n] \tag{2}$$

The final step to calculate the envelope signal is to take the magnitude value of the complex analytic signal, as described in Equation (3).

$$EA[n] = (x_r^2[n] + x_i^2[n])^{\frac{1}{2}} \tag{3}$$

where $EA[n]$ represents the envelope signal.

3.2. Feature Compression

Feature compression was applied to both the discrete-time STFT- and EA-generated spectral feature sets referenced in Sections 3.1.1 and 3.1.2. For the STFT features, dimensionality reduction from 512 down to 25 spectral features was applied in order to extract the most useful features to train the learning model algorithm. In addition, dimensionality reduction simplifies the complexity of the computations so that more optimal and accurate estimations could be obtained. This was to avoid the well-known phenomena often defined as the *curse of dimensionality* [70,71]. This term describes the inherent problem caused by

the exponential increase in volume associated with adding extra dimensions to a Euclidean space [71].

Feature reduction was achieved using a filter band approach as described in Equation (4) for both linear and non-linear sized bands. The linear bands, $L[m,k]$, consisted of equidistant bands applied evenly to 512 spectral features, whereas the non-linear approach, $O[m,k]$, comprises a one-third octave band filter being applied. The one-third octave band filter places a higher emphasis on the lower end of the frequency spectrum by having smaller bands that increase non-linearly in size as the frequency increases.

$$Y[m,k] = X[m,f] \times H[f,k] \tag{4}$$

where the feature vector $Y[m,k]$ is the result of a matrix multiplication of $X[m,f]$, the average STFT values of each discrete time-sample, and $H[f,k]$, a matrix comprised of ones and zeros strategically placed according to the desired filter parameters. The feature vector $Y[m,k]$ is equal to $L[m,k]$ when linear filtering is used, and $Y[m,k]$ is equal to $O[m,k]$ when one-third octave band filtering is used.

Two distinct $H[f,k]$ matrices were chosen and tested, allowing for linear and non-linear banding. The linear band feature reduction technique was achieved by filtering the frequency amplitude spectrum of each independent time-sample into 25 equidistant bands ranging from 0 to 12,800 Hz. The non-linear feature reduction approach uses a one-third octave band filter scale to segment the frequency spectrum into 25 octave bands in the scale 0 to 12,800 Hz.

For the signal envelope features, $EA[n]$, three feature compression techniques were explored using a linear and one-third octave band scale. The first involves the time-series values for each of the 25 filter bands being transferred to the frequency domain by means of the Fast Fourier Transform (FFT). The sum of the frequency powers, in decibel scale, is calculated to create feature vectors comprising of 25 features for each of the discrete-time vibration signals, $\mathcal{F}\left\{EA_L\right\}$ and $\mathcal{F}\left\{EA_O\right\}$. The second technique is comprised of summing all the signal envelope time-series points for each of the 25 filter band vectors, EA_L and EA_O. The third approach used the signal energy value by calculating the square of the signal envelope values and summing together for each of the 25 filter band, EA_L^2 and EA_O^2.

3.3. Wear-State Temporal Models

When analysing the horizontal vibration signal from the bearing's external race, a degradation trend can be identified in the signal amplitude values in the discrete-time domain. The degradation trend can be identified as having a non-linear increase in amplitude as the bearing failure stage is approached. In this study, five temporal wear-state classes were used to characterise the RUL of the seven bearings under test.

Two different wear-state models, one *linear* and one *non-linear*, were considered and tested to determine the optimum scale to determine the RUL of the components accurately and robustly. The linear wear-state model divides the data into five equidistant temporal classes, as illustrated by Equation (5), where each temporal class represents a 20% portion of the bearing's overall lifetime. In contrast, the non-linear wear-state approach uses five classes that are strategically spaced to add greater granularity or compression to the class boundaries towards the latter stages of the bearing's life. This is achieved as follows: the first 63% of the bearing's lifetime is allocated to class 1 (healthy), the second class consists of the 86% of the bearing's lifetime and so on as presented in Equation (6).

$$\alpha_i = \left(\frac{1}{5}\right) \times i \tag{5}$$

$$\beta_i = 1 - e^{-i} \tag{6}$$

where α_i and β_i define the linear and non-linear temporal class boundaries, respectively, and the index $i = \{1,2,3,4\}$ corresponds to those class numbers. Note: $\alpha_5 = \beta_5 = 1$ as the

boundaries are normalised with respect to time.

Figure 5 shows a graphical representation of the linear and non-linear wear-state class boundaries.

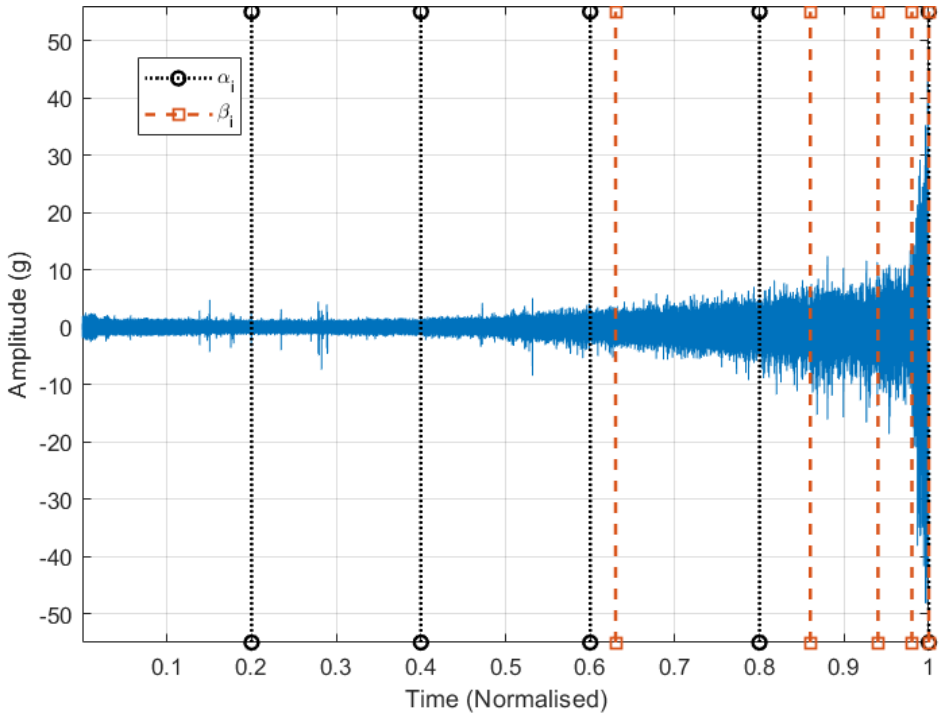


Figure 5. Linear and non-linear wear-state class boundaries, α_i and β_i, respectively.

3.4. Classification

Supervised ML algorithms were used to detect trends and patterns in the pre-processed data and classify the health wear-state of the bearing test samples. Two widely used classical methods of ML were studied—that of support vector machines (SVM) and *k*-Nearest Neighbour (*k*-NN).

3.4.1. Support Vector Machines

A support vector machine (SVM) is a supervised ML algorithm. The SVM algorithm is used to classify pre-labelled test cases (targets) by analysing the training cases (predictors) and finding a separator between classes. The use of SVM-classification algorithms has been recorded to achieve highly accurate results in high dimensional feature spaces [52,53,55,72,73]. Another key benefit that the SVM algorithm option offers is memory efficiency. Only a small subset of the training features, the support vectors, are required to calculate the location of optimised hyperplanes between wear-state classes.

The pre-labelled training data, also referred to as the predictor features, is mapped to a higher-dimensional feature space, so that data points can be categorised as shown in Figure 6. This mapping process is often referred to as kernelling, as the transformation can be achieved through the application of various kernel functions. The predictor features are transformed in such a way that the separator can be formed as a hyperplane, which can be considered as a line function representing the largest separation, or margin, between classes (wear-states). The data points whose positions lie closest to the calculated hyperplane are identified as support vectors. To achieve the most accurate SVM prediction model, the hyperplane should be at the maximal distance possible from the nearest support vectors.

This distance from a support vector to the hyperplane is identified as the margin. The classification of target instances is achieved by inputting the unseen, feature data

values, without their corresponding wear state label, into the hyperplane function shown in Equation (7).

$$w^T y + b = 0 \tag{7}$$

where w represents the weight vector, y is the input vector and b is the bias. The result determines whether the data point is an instance of the class above or below the hyperplane.

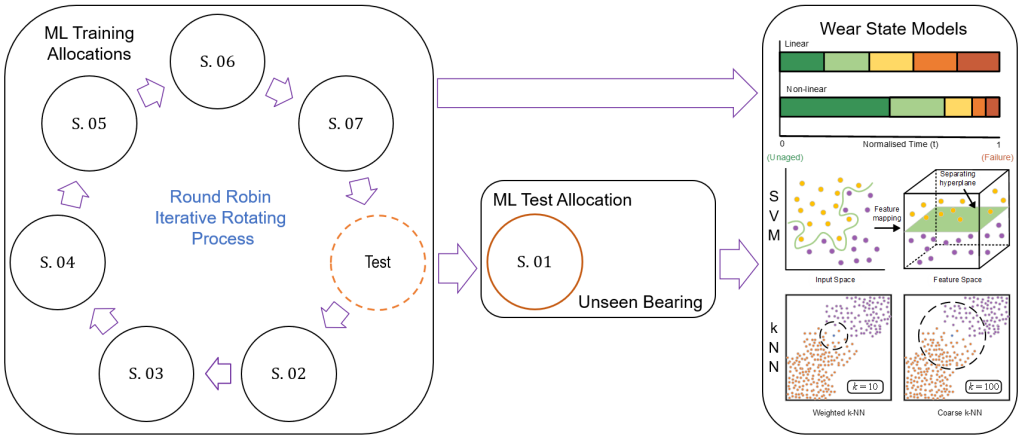


Figure 6. Experimental procedure for selecting the training and testing signals for the different ML approaches.

This means that only a fraction of the overall predictor feature points are processed for calculation unlike other ML methods, such as Decision Trees, Logistic Regression, Naive Bayes and *k*-NN classification, which require all feature points to be included in each calculation. This significantly reduces the complexity of calculations while increasing the efficiency and speed of producing RUL estimations. The biggest challenge associated with the SVM classification algorithm as a prediction model is its tendency to over-fit data. Over-fitting would be most prevalent when the feature number (dimensionality) is high relative to the number of predictor instances. To counteract this, numerous bearing test-cases with large numbers of time-samples are used to train and test the performance of the SVM classification algorithms.

The performance of six different kernel functions were investigated and compared in this study. Kernel options used to map the data into a higher dimensional feature space included Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian and Coarse Gaussian functions.

3.4.2. *k*-Nearest Neighbour

The *k*-Nearest Neighbour (*k*-NN) classification algorithm is one of the most widely used supervised ML methods for categorising unknown signals into a discrete set of classes [49–51]. The *k*-NN method is used to classify target instances based on their similarities to predictor features (training data). The most similar predictor cases are referred to as the "neighbours", hence, the title associated with the classification method.

Classification is achieved by first defining a value of *k*. The optimal *k* value is dependant on the input data. Choosing a low *k* value often produces an over-fitted prediction model, which produces inaccurate predictions on out-of-sample target instances. A higher *k* value makes the prediction model too generalised as the classes with more predictor instances become prioritised as the target instance. The optimal *k* value may only be determined through trial and error, using multiple values to compare accuracy results. The next step to achieve a RUL prediction using the *k*-NN framework involves calculating the distance from the features of the target instance from all other predictor instance features. This distance metric can be calculated in a number of ways, including Euclidean, Mahalanobis, City block and Minkowski distances.

As we are dealing with distance metrics to determine classification, it is important that each of the features used are standardised before training the prediction model. Min–max normalisation is applied to each feature to put all training data into the 0–1 scale. The target instance is then normalised using the same min–max values as the training data.

The same min–max value is used for the normalization process to eliminate the occurrence of data dredging, the statistical inference performed after looking at the complete dataset. The k nearest observations in the training data that are nearest to the unknown target data point are selected as the "neighbour" points. A case is classified by determining the mode class value of its neighbours, with the case being assigned to the class most common amongst its k nearest neighbours measured by a distance function.

Six methods of k-NN classification were used to obtain RUL predictions for the bearings, including Fine, Medium, Coarse, Cosine, Cubic and Weighted k-NN. The specified parameters varied in each method are presented in Table 3.

Table 3. k-NN method parameters.

k-NN Method	Distance Weightings	kValue
(Fine)	Euclidean	1
(Medium)	Euclidean	10
(Coarse)	Euclidean	100
(Cosine)	Cosine	10
(Cubic)	Cubic	10
(Weighted)	Weighted	10

4. Experimental Procedure

This section describes the experimental procedure, which can be summarised under three main strands: the ML method recipes, the round robin framework and the performance metrics.

4.1. ML Method Recipes

The experimental procedure for this research involved varying the following parameters as described in the previous section: (a) feature extraction using either STFT of the discrete-time domain signal or the envelope of the vibration time-series data, (b) feature selection using full spectra from 0 to 12,800 Hz $X[m, f]$, linear bands $L[m, k]$ or one-third octave bands $O[m, k]$ as feature vectors, (c) the wear-state classification model using either linear, α, or non-linear, β, temporal class boundaries, (d) model training and testing using a SVM or k-NN method approach. In the case of SVM kernelling six function options including: linear, quadratic, cubic, fine, medium and coarse Gaussian, were applied to convert the input signals to a higher dimensional feature space. Six classification methods for determining the target class for the k-NN algorithm were investigated including fine, medium, coarse, cosine, cubic and weighted.

4.2. Round Robin Framework

All seven bearing test cases were incorporated into each RUL estimation process in a round robin framework that seeks to maximise the data set as well as mitigate problems relating to over-fitting. The experimentation process involved allocating six bearing signals as training datasets to teach the ML algorithms. The seventh bearing test-case was used for testing purposes. Once RUL estimations were obtained for each of the out-of-sample test signals, the bearing was added back to the in-sample testing pool, and the next sequential bearing was transferred to the testing pool. The ML prediction model was retrained, and this was iterated until RUL estimations had been obtained for all seven bearing test cases.

The incorporation of this framework to train and test the performance of each prediction algorithm greatly reduces over-fitting as we are only using out-of-sample test signals. All model training data comprises of signals from a completely different bearing for each test case. This gives an extremely accurate interpretation of how the models would perform

in a real-world application using signals from different bearings used to train the models in every case.

The classification process involves dealing with a multi-class and multi-label classification model, which comprises five temporal wear-state classes to be estimated for thousands of consecutive time samples. A moving-average (MA) filtering technique is incorporated to smooth out any undesirable whipsawing or erratic transitioning between the five temporal wear-state classes. The MA technique involves taking a window length of nine discrete-time samples, consisting of the current target prediction, the previous four predicted targets and the following four predictions. The mode of the nine predicted values is then assigned to the current target sample. In a real-time application, these nine samples comprise a 40 s temporal time period. This short time period is extremely negligible over a bearing's lifetime, which is typically years for a real system.

4.3. Performance Metrics

The performance of each ML approach investigated was analysed by computing the Jaccard Index [74,75], Equation (8) and multiplying by a factor of 100 to obtain a percentage accuracy value.

$$J(z, \hat{z}) = \frac{|z \cap \hat{z}|}{|z \cup \hat{z}|} \tag{8}$$

where z represents the true class of a time-sample and $\hat{z}$ represents the class prediction from the ML algorithm.

The Mean Absolute Error (MAE) was calculated for each of the classification models and feature selection options [76,77]. Using Equation (9), the absolute error between the predicted target instances and the real expected values was calculated for each time-sample. This resulted in a natural number in the range 1 to 4, as we are dealing with five wear-state classes and the maximum error a prediction could possibly be classified as would be four classes away from the expected real class value. The error for each individual time-sample was summed and divided by the total number of test-samples to calculate the MAE. These MAE values for each classification model were then normalised by dividing each MAE result by 4 and are compared in the tables below.

$$MAE = \frac{1}{M} \sum_{m=1}^{M} |z_m - \hat{z}_m| \tag{9}$$

where M represents the total number of target instances to be classified, z represents the true class of a target instance and $\hat{z}$ represents the class prediction from the ML algorithm.

5. Results

This section presents the results obtained from the proposed ML framework for RUL classification.

5.1. Linear Wear-State Classification Approach

The linear wear-state classification accuracy and MAE error results achieved using the STFT features extracted from the discrete-time signals are presented in Table 4. In the case of the SVM classification method, the lowest accuracy results were recorded at 27.5% for the one-third octave band feature set using a fine Gaussian kernel function. The highest classification performance achieved was 59.5% using the one-third octave band feature selection and a coarse Gaussian kernelling method.

For the k-NN classification method, the lowest classification performance recorded was 39.9% for the linear band feature set using fine k-NN. The highest classification accuracy achieved was 54.2% by the one-third octave band features with cosine k-NN. The MAE results indicate that the coarse Gaussian kernel using the one-third octave band features was also the best-performing model with the lowest MAE score. The lowest error scores of 0.17 were achieved by both the cubic and cosine k-NN models.

The results of the experiment using the signal envelope-derived features and linear wear-state classification are presented in Table 5. In the case of the SVM model, the lowest performance was recorded at 30.4% accuracy for the one-third octave band FFT features using a fine Gaussian SVM classifier. The highest classification performance, on the other hand, was achieved at 62.5%, for the one-third octave band summed envelope features using a Linear SVM kernel function.

For the *k*-NN classification results, the lowest performance was recorded at 44.2% accuracy for the squared one-third octave band features from a fine *k*-NN classifier. The highest classification accuracy was achieved at 57.9% for the squared linear band features using a coarse *k*-NN classifier. The MAE results indicate that the linear, coarse and medium Gaussian kernel functions using the summed one-third octave band features were the best-performing model with the lowest MAE score. The coarse *k*-NN model proved to be the best option from the MAE values also, proving to be the most accurate in both Jaccard Index and MAE aspects.

Table 4. Linear wear-state classification results (%) and MAE (normalised) for STFT feature sets, where G denotes Gaussian.

			***k*-NN**			
Features	**Fine**	**Medium**	**Coarse**	**Cosine**	**Cubic**	**Weighted**
$\mathcal{F}\left\{x[n]\right\}$	53.7 | 0.16	48.5 | 0.17	48.0 | 0.17	32.3 | 0.39	50.4 | 0.16	56.7 | 0.14
$L[m,k]$	48.7 | 0.18	50.6 | 0.17	43.7 | 0.22	33.1 | 0.36	48.2 | 0.17	55.1 | 0.14
$O[m,k]$	57.1 | 0.15	51.2 | 0.17	51.1 | 0.17	27.5 | 0.40	56.8 | 0.14	**59.5** | **0.13**
			***k*-NN**			
Features	**Fine**	**Medium**	**Coarse**	**Cosine**	**Cubic**	**Weighted**
$\mathcal{F}\left\{x[n]\right\}$	41.5 | 0.23	43.8 | 0.22	44.7 | 0.19	45.1 | 0.20	43.7 | 0.21	43.0 | 0.22
$L[m,k]$	39.9 | 0.23	41.6 | 0.23	45.0 | 0.19	45.0 | 0.21	43.9 | 0.22	41.7 | 0.23
$O[m,k]$	52.2 | 0.19	51.5 | 0.18	53.4 | 0.17	**54.2** | **0.17**	50.7 | 0.19	51.8 | 0.18

Table 5. Linear wear-state classification results (%) and normalised MAE for envelope analysis (EA), where G denotes Gaussian.

			SVM			
Features	**Linear**	**Quadratic**	**Cubic**	**Fine G**	**Medium G**	**Coarse G**
$\mathcal{F}\left\{EA_L\right\}$	47.9 | 0.20	48.5 | 0.18	42.0 | 0.22	36.9 | 0.32	51.9 | 0.17	51.2 | 0.16
EA_L	55.3 | 0.15	50.6 | 0.19	37.2 | 0.28	46.8 | 0.23	61.8 | 0.12	59.9 | 0.12
EA_L^2	47.3 | 0.20	32.9 | 0.32	42.0 | 0.23	47.6 | 0.18	56.9 | 0.14	54.2 | 0.15
$\mathcal{F}\left\{EA_O\right\}$	56.1 | 0.15	50.9 | 0.17	54.8 | 0.15	30.4 | 0.38	60.4 | 0.13	53.9 | 0.14
EA_O	**62.5** | **0.12**	54.7 | 0.16	50.9 | 0.18	41.6 | 0.28	56.7 | 0.15	58.3 | 0.12
EA_O^2	53.6 | 0.17	37.9 | 0.28	32.0 | 0.31	37.1 | 0.23	59.9 | 0.14	53.6 | 0.15
			***k*-NN**			
Features	**Fine**	**Medium**	**Coarse**	**Cosine**	**Cubic**	**Weighted**
$\mathcal{F}\left\{EA_L\right\}$	46.9 | 0.20	47.2 | 0.20	46.0 | 0.19	46.8 | 0.22	47.0 | 0.20	47.5 | 0.20
EA_L	51.8 | 0.18	52.9 | 0.18	55.6 | 0.15	50.7 | 0.18	53.0 | 0.17	53.0 | 0.18
EA_L^2	55.6 | 0.15	57.0 | 0.15	**57.9** | **0.14**	55.5 | 0.15	56.7 | 0.15	56.1 | 0.15
$\mathcal{F}\left\{EA_O\right\}$	53.4 | 0.18	51.8 | 0.17	51.0 | 0.17	51.0 | 0.17	51.3 | 0.18	52.2 | 0.17
EA_O	48.4 | 0.22	49.4 | 0.21	46.9 | 0.22	47.4 | 0.22	48.6 | 0.22	49.7 | 0.21
EA_O^2	44.2 | 0.25	44.6 | 0.24	43.4 | 0.25	42.6 | 0.25	44.2 | 0.26	44.4 | 0.25

5.2. Non-Linear Wear-State Classification Approach

The results achieved using non-linear wear-state classes and STFT features are presented in Table 6. For the SVM experiments, the lowest performance recorded was 53.9% for the linear band features using a cubic kernel function SVM model. The highest performance of 73.6% was achieved by the one-third octave band features using a medium Gaussian kernel function.

In the case of the *k*-NN experiments, the lowest performance recorded was 60.1% from the linear frequency band features with the Fine *k*-NN. The highest performance accuracy of 73.2% was achieved using the one-third octave band features with Coarse *k*-NN. This was also the highest classification performance achieved overall for the STFT analysis feature study.

The normalised MAE results for the STFT spectral features using SVM classification models indicate that the Medium and Coarse Gaussian kernel function using the one-third octave band features was also the best-performing model with the lowest normalised MAE score. Similarly for the *k*-NN experiments, the one-third octave using Coarse *k*-NN achieved the lowest error score.

Table 6. Non-linear wear-state classification results (%) and normalised MAE for STFT data, where G denotes Gaussian.

			SVM			
Features	**Linear**	**Quadratic**	**Cubic**	**Fine G**	**Medium G**	**Coarse G**
$\mathcal{F}\left\{x[n]\right\}$	66.8 | 0.12	61.6 | 0.18	59.1 | 0.19	63.2 | 0.14	65.6 | 0.13	65.4 | 0.11
$L[m,k]$	59.1 | 0.15	59.7 | 0.19	53.9 | 0.23	63.2 | 0.15	60.4 | 0.15	65.1 | 0.11
$O[m,k]$	71.2 | 0.11	68.8 | 0.15	62.8 | 0.17	62.4 | 0.15	**73.6** | **0.10**	68.4 | 0.10
			***k*-NN**			
Features	**Fine**	**Medium**	**Coarse**	**Cosine**	**Cubic**	**Weighted**
$\mathcal{F}\left\{x[n]\right\}$	64.0 | 0.16	65.0 | 0.15	66.7 | 0.11	66.9 | 0.15	64.8 | 0.15	64.5 | 0.15
$L[m,k]$	60.1 | 0.18	62.0 | 0.17	64.8 | 0.11	63.2 | 0.15	63.2 | 0.17	61.0 | 0.17
$O[m,k]$	69.4 | 0.13	68.9 | 0.12	**73.2** | **0.10**	69.5 | 0.12	68.7 | 0.11	69.7 | 0.12

The non-linear wear-state results for the Signal-Envelope-derived features using are presented in Table 7. For the SVM models, the lowest performance recorded was 9.5% for the squared one-third octave band features using a cubic kernel function, whereas the highest performance of 73.1% was achieved for the summed one-third octave band envelope features using a linear SVM kernel function. In the case of the *k*-NN classification results, the lowest performance recorded was 62.0% for the squared one-third octave band features using a Fine *k*-NN classifier.

The highest performance accuracy was 74.3% for the one-third octave band FFT features using a cosine *k*-NN classifier. This was also the highest classification performance achieved overall for the Envelope Analysis *k*-NN study and importantly for all of the experimental studies reported in this work. The normalised MAE results using the signal envelope derived spectral features applying SVM classification models indicate that the Linear kernel function using the one-third octave band features were the best-performing model with the lowest MAE score of 0.09.

The Cosine *k*-NN model proved to be the best options for the non-linear temporal classes, with an error score of 0.08, which was also the lowest overall error score across all four experimental studies presented. Accordingly, these best MAE values also correspond with the best classification accuracy achieved, which is not unexpected.

Table 7. Non-linear wear-state classification results (%) and normalised MAE for envelope analysis features, where G denotes Gaussian.

			SVM			
Features	**Linear**	**Quadratic**	**Cubic**	**Fine G**	**Medium G**	**Coarse G**
$\mathcal{F}\{EA_L\}$	63.7 | 0.13	63.5 | 0.17	60.4 | 0.19	63.9 | 0.15	67.8 | 0.12	63.3 | 0.12
EA_L	65.9 | 0.11	64.8 | 0.16	57.8 | 0.19	66.0 | 0.16	72.0 | 0.09	68.5 | 0.11
EA_L^2	62.7 | 0.13	14.7 | 0.68	27.7 | 0.38	66.9 | 0.13	69.1 | 0.11	68.1 | 0.12
$\mathcal{F}\{EA_O\}$	71.2 | 0.20	65.5 | 0.30	67.5 | 0.31	65.2 | 0.27	73.0 | 0.21	68.2 | 0.21
EA_O	**73.1** | **0.09**	68.1 | 0.13	56.7 | 0.23	60.4 | 0.18	66.3 | 0.11	68.9 | 0.12
EA_O^2	66.7 | 0.11	36.2 | 0.41	9.5 | 0.48	60.9 | 0.15	70.2 | 0.11	70.3 | 0.11
			***k*-NN**			
Features	**Fine**	**Medium**	**Coarse**	**Cosine**	**Cubic**	**Weighted**
$\mathcal{F}\{EA_L\}$	67.5 | 0.15	65.2 | 0.15	68.3 | 0.13	64.1 | 0.16	65.4 | 0.15	65.0 | 0.15
EA_L	69.8 | 0.14	71.6 | 0.13	72.5 | 0.10	70.7 | 0.13	71.8 | 0.13	71.7 | 0.13
EA_L^2	73.1 | 0.12	73.0 | 0.11	73.5 | 0.10	72.4 | 0.10	73.2 | 0.11	72.8 | 0.11
$\mathcal{F}\{EA_O\}$	74.1 | 0.09	73.0 | 0.08	72.0 | 0.08	**74.3** | **0.08**	72.0 | 0.08	73.3 | 0.08
EA_O	66.0 | 0.15	64.4 | 0.13	67.0 | 0.12	63.0 | 0.14	63.7 | 0.13	64.4 | 0.13
EA_O^2	62.0 | 0.16	63.3 | 0.14	63.1 | 0.13	66.0 | 0.15	63.1 | 0.15	63.0 | 0.15

6. Discussion

The results presented in Section 5 highlight the best- and worst-performing supervised ML approaches for both the linear and the non-linear wear-state classes that were investigated. The SVM-based algorithm approach that employed one-third octave band-based features was found to yield the best performance for the linear wear-state classes. This was the case for both STFT- and EA-based features, achieving scores of J = 59.5% with MAE = 0.13 for $O[m,k]$ using SVM (Coarse G) and J = 62.5% with MAE = 0.12 for EA_O using SVM (Linear), respectively, as shown in Tables 4 and 5, for the linear wear-state classification.

Again, for the non-linear wear-state classification the SVM (Medium G) based algorithm performed extremely well using the $O[m,k]$ STFT features by achieving scores of J = 73.6% with MAE = 0.10 as shown in Table 6. Using the same octave band feature set, the *k*-NN (Cubic) approach had comparable performance coming in at slightly less accuracy at J = 73.2% with the same MAE = 0.10, as shown in Table 6. In the case of the EA features for non-linear wear-state classification, the SVM (Linear) achieved J = 73.1% with MAE = 0.09, see as shown in Table 7. However, the *k*-NN (Cosine) had superior performance using the spectral-based EA features, $\mathcal{F}\{EA_O\}$, achieving J = 74.3% with MAE = 0.08. This was the best performance achieved for all the ML approaches over this entire investigation, with the best classification accuracy and the lowest MAE.

In order to better analyse and interpret the results more closely, confusion matrices are presented for the best-performing approaches for both the linear and non-linear wear-state classification approaches investigated. Accordingly, these confusion matrices correspond to the approaches that have their values highlighted in bold font in Tables 4–7 as discussed previously. At the class level, these confusion matrices enable the classification results to be examined, and they allow the MAE to be quantified and better appreciated—for instance, regarding how many samples from class 1 were incorrectly classified as class 5.

The confusion matrices shown in Figure 7 allow us to see the classification performance for each class by observing the percentage score along the diagonal. It is noted that the vast majority of classification inaccuracy (MAE) tends to be predicting the neighbouring class, which is significant, while this information was captured collectively for the entire class set using the MAE metric for ML approaches; however, these confusion matrices offer the granularity to identify which specific classes were the most challenging to estimate.

All of the ML recipes presented performed very well on class 1, the max and min range being 92.6% to 89.0% in comparison to the max and min range for class 5 of 68.1% to 41.6%. The ML recipe with the best performance overall at 62.5% with MAE of 0.12 for the linear wear-state classification was that of the EA_O features with a SVM (Linear) algorithm, this is shown in matrix (c) of Figure 7. This particular ML recipe achieved the highest classification for class 4 at 65.1% and was second best in class 1, 3 and 5, which ultimately led to it achieving the best overall score.

Similarly, as shown in Figure 8, the classification performance for class 1 for the non-linear wear-state classes was good; however, the range was wider, with max and min values of 95.5% and 78.9%, respectively. Whereas in class 5, the max and min range was from 63.2% to 39.3%. The best performance overall at 74.3% with MAE of 0.08 for the linear wear-state classification was the ML recipe that comprised of the EA_L^2 features with k-NN (Coarse) algorithm, this is shown in matrix (d) of Figure 8. This ML recipe strong performance in class 1 and average performance in class 2 and 3 with poor performance in class 5.

However, the performance in class 4 at 25.7% significantly outperformed the other ML recipes shown, and this lead to it scoring the best overall. These trends in individual class performance also be viewed in Figure 9, without the benefit of visualising where the incorrectly classified test cases have been predicted. These points along with a mean value correspond to the diagonal values for the confusion matrices in Figures 7 and 8. The high performance of class 1 for both the linear and non-linear wear-state class options is identifiable as well as the decreasing trend as the classes approach the failure stage of the bearings under test.

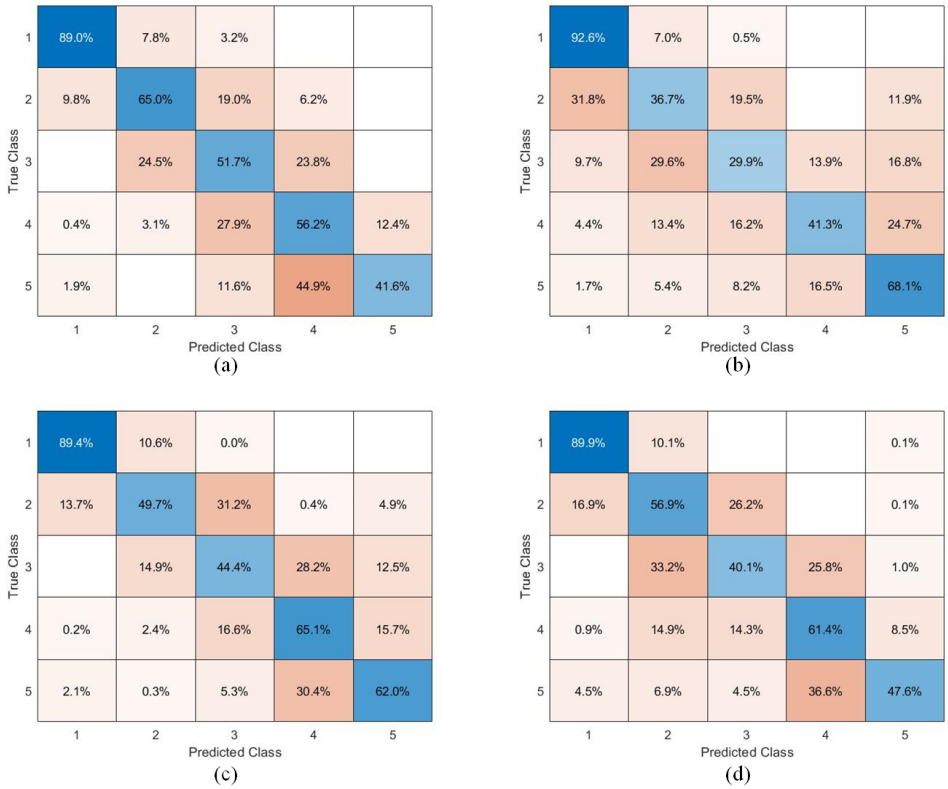


Figure 7. Confusion matrices of the best-performing ML recipes for the linear wear-state classes: (**a**) $O[m,k]$ with SVM (Coarse G), (**b**) $O[m,k]$ with k-NN (Cosine), (**c**) EA_O with SVM (Linear) and (**d**) EA_L^2 with k-NN (Coarse).

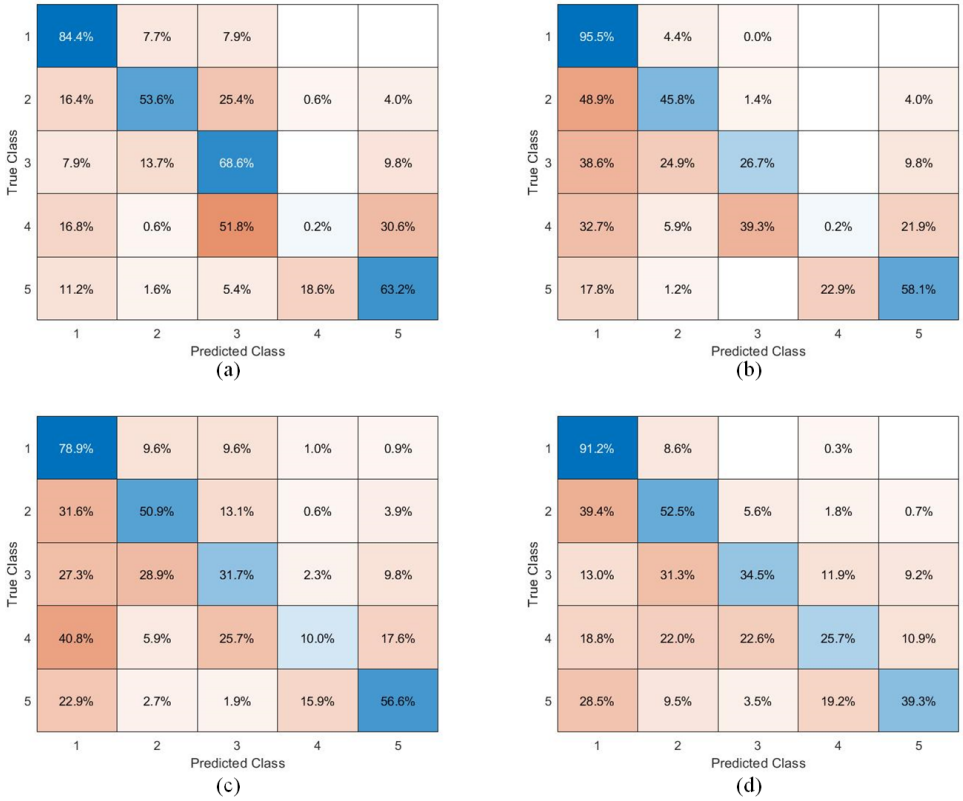


Figure 8. Confusion matrices of the best-performing ML recipes for the non-linear wear state classes: (**a**) $O[m,k]$ with SVM (Medium G), (**b**) $O[m,k]$ with k-NN (Coarse), (**c**) EA_O with SVM (Linear) and (**d**) $\mathcal{F}\left\{EA_O\right\}$ with k-NN (Cosine).

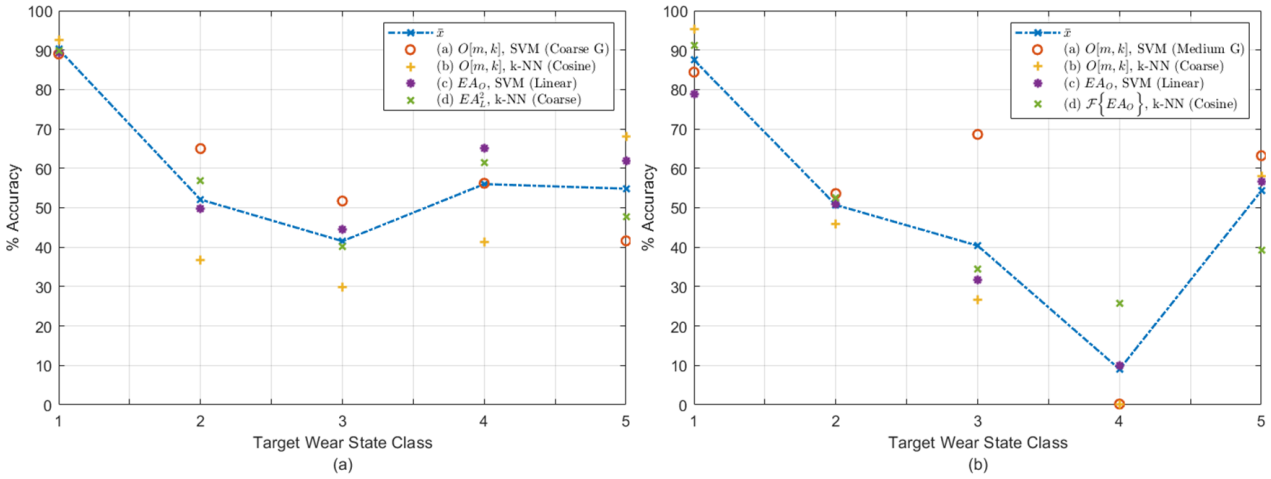


Figure 9. Classification accuracy for the individual classes (1 to 5) for (**a**) linear wear-state classes and (**b**) non-linear wear-state classes, which corresponds to the diagonal of the confusion matrices in Figures 7 and 8, respectively.

Prior work by Sutrisno et al. [78], Singleton et al. [79]a nd Lei et al. [80], presented ML methods that achieved percentage accuracy scores of 76.2%, 67.20% and 77.44%, respectively, using the PRONOSTIA bearing dataset. However, these proposed methods utilised a framework where only bearings S. 01 and S. 02 were used for training the algorithm, and the remaining five bearings were used for testing. The round-robin experimental framework presented in this paper presents the mean percentage accuracy of all seven bearing signals whereas the prior work only presented the mean of five signals. Furthermore,

the MAE performance metric was used for analysis purposes to ascertain the severity of the misclassifications.

Feature subset compression using the non-linear one-third octave-based filtered for both linear and non-linear wear-states performed very well in both cases. This can be attributed to placing a higher emphasis on the lower portion of the spectra for feature extraction. From a feature-engineering perspective, this was shown to offer more valuable diagnostic trend information for characterising the health condition of the bearings under test [11].

Importantly, this reduces the multivariate dimensionality [70,71] of the feature space in a more optimal way compared with linear filtering as the results demonstrated by yielding superior classification performance. Moreover, using a non-linear wear-state model approach to classification is more suitable as the ageing mechanisms typically follow an non-linear exponential trend; hence, we can see significantly higher performances achieved. Clearly, a trade-off exists as if the size of class 1, is too large with respect to the others, then the suitability of the RUL framework for taking timely action, such as equipment maintenance and critical parts replacement diminishes.

As the subsequent classes would therefore be too short in time, the severity of degradation between these classes would occur rapidly. Our non-linear exponential model described in Equation (6) strikes a suitable balance and was found to work extremely well in this proposed approach.

The highest overall classification scores were achieved using the Cosine *k*-NN classifier. This was achieved across all seven bearing test-cases using the round robin framework and, hence, demonstrates how the proposed ML methodology performs on unseen raw vibration signals. However, these *k*-NN and SVM ML algorithms are heavily reliant on the depth of the training data, which is common in the field of supervised learning and hence makes these methods prone to producing over-fitted prediction models. While this paper introduced a valuable and robust approach for RUL estimation, future work might investigate applying this proposed method on larger data-sets.

The dataset used here in this research was limited in terms of the total number specimens aged and captured using vibration signals In addition, the level of accelerated ageing is perhaps too rapid, which led to a high proportion of abrupt failure modes occurring approximately 43% of the time. These have a completely different degradation trend to the gradual ageing mechanisms; hence, this places limits on testing the true efficacy of the ML frameworks and recipes due to model over-fitting. If the datasets were significantly improved by increasing the number of specimens and reducing the level of accelerated ageing, this would offer the potential to explore ML approaches that employ advanced deep learning using neural networks.

In testing the versatility and robustness of the proposed ML method recipes on different bearing types and sizes under different speed and load conditions, work could explore vibration data gathered from research testbeds where the shaft speed changes. This will require developing extensive experimental campaigns to create more advanced datasets that better reflect typical real-world operating conditions.

7. Conclusions

Traditionally, condition-based monitoring (CbM) of electric and rotating machines has focused heavily on two primary areas, the *detection* and the *diagnosis* of fault modes. More recently, research efforts have investigated the more challenging area of *prognosis* to determine the remaining useful life (RUL) of the machine under test. This paper introduced a valuable machine learning (ML) approach to estimate the RUL of rolling element bearings, which are a core component of rotating machines.

The proposed ML recipes and approaches comprise of signal processing techniques and ML algorithms applied to real-world vibration signals, which were acquired from the outer-race of bearings degraded over time using an accelerated ageing test-rig. The paper reports the results for linear and non-linear wear-state models using novel feature

engineering derived from Short-Time Fourier Transform (STFT) and Envelope Analysis (EA) representations. In addition, two different classification algorithm approaches, *k*-Nearest Neighbour (*k*-NN) and Support Vector Machines (SVM), were investigated and compared.

This work achieved classification accuracy results of up to 74.3% with a mean absolute error (MAE) of 0.08, which demonstrates the method's efficacy for performing the task of RUL. This ultimately offers a robust and low complexity approach that is highly valuable for advanced predictive maintenance purposes in industry.

Author Contributions: The following section details the contributions made by each author to the research that was presented in this journal article submission: E.T.C.: Formal analysis, Methodology, Investigation, Writing—Original Draft and Writing—Review and Editing; V.I.M.: Supervision, Analysis, Results interpretation, Writing—Review and Editing and Visualization; D.F.K.: Supervision, Formal analysis, Methodology, Analysis, Results interpretation, Writing—Review and Editing, Plotting and Visualization. All authors have read and agreed to the published version of the manuscript.

Funding: The research conducted in this publication was funded by the Irish Research Council under award number GOIPG/2021/1744. All authors are with the Faculty of Engineering, Institute of Technology Carlow, Ireland.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analysed in this study. This data can be found here: (https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#bearing, accessed on 11 March 2022), also see [62].

Conflicts of Interest: The authors declare no conflict of interest. The Institute of Technology Carlow have endorsed this manuscript to go forward for peer review and publication. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ADC	Analogue-to-digital converter
ANN	Artificial neural network
CBM	Condition-based monitoring
DFT	Discrete Fourier transform
EA	Envelope analysis
EMD	Empirical mode decomposition
FFT	Fast Fourier transform
FIR	Finite impulse response
k-NN	*k*-Nearest neighbour
LSTM	Long short-term memory
MA	Moving averaging
MAE	Mean absolute error
MCSA	Motor Current signal analysis
ML	Machine learning
RUL	Remaining useful life
STFT	Short-Time Fourier transform
SVM	Support vector machine
WT	Wavelet transform

References

1. Kudelina, K.; Vaimann, T.; Asad, B.; Rassõlkin, A.; Kallaste, A.; Demidova, G. Trends and Challenges in Intelligent Condition Monitoring of Electrical Machines Using Machine Learning. *Appl. Sci.* **2021**, *11*, 2761. [CrossRef]
2. Tavner, P.J. Review of condition monitoring of rotating electrical machines. *IET Electr. Power Appl.* **2008**, *2*, 215–247. [CrossRef]
3. Xu, L.; Chatterton, S.; Pennacchi, P. Rolling element bearing diagnosis based on singular value decomposition and composite squared envelope spectrum. *Mech. Syst. Signal Process.* **2021**, *148*, 107174. [CrossRef]

4. Kulkarni, V.V. Condition monitoring analysis of rolling element bearing based on frequency envelope. *Mater. Today Proc.* **2021**, *46*, 4667–4671. [CrossRef]
5. Nirwan, N.W.; Ramani, H.B. Condition monitoring and fault detection in roller bearing used in rolling mill by acoustic emission and vibration analysis. *Mater. Today Proc.* **2021**, *51*, 344–354. [CrossRef]
6. Gyftakis, K.N.; Cardoso, A.J.M. Reliable Detection of Very Low Severity Level Stator Inter-Turn Faults in Induction Motors. In Proceedings of the IECON 2019–45th Annual Conference of the IEEE Industrial Electronics Society, Lisbon, Portugal, 14–17 October2019; Volume 1, pp. 1290–1295.
7. Zhang, P.; Du, Y.; Habetler, T.G.; Lu, B. A Survey of Condition Monitoring and Protection Methods for Medium-Voltage Induction Motors. *IEEE Trans. Ind. Appl.* **2011**, *47*, 34–46. [CrossRef]
8. Nandi, S.; Toliyat, H.A.; Li, X. Condition Monitoring and Fault Diagnosis of Electrical Motors—A Review. *IEEE Trans. Energy Convers.* **2005**, *20*, 719–729. [CrossRef]
9. Singh, G.; Kazzaz, S.A.S.A. Induction machine drive condition monitoring and diagnostic research—A survey. *Electr. Power Syst. Res.* **2003**, *64*, 145–158. [CrossRef]
10. Saidi, L.; Benbouzid, M. Prognostics and Health Management of Renewable Energy Systems: State of the Art Review, Challenges, and Trends. *Electronics* **2021**, *10*, 2732. [CrossRef]
11. Chelmiah, E.T.; McLoone, V.I.; Kavanagh, D.F. Remaining Useful Life Estimation of Rotating Machines using Octave Spectral Features. In Proceedings of the IECON 2020 the 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 18–21 October 2020; pp. 3031–3036.
12. Lebold, M.; Thurston, M. Open standards for condition-based maintenance and prognostic systems. In Processing of the 5th Maintenance and Reliability Conference (MARCON 2001), Gatlinburg, TN, USA, 6–9 May 2001; Volume 200.
13. Xu, L.; Pennacchi, P.; Chatterton, S. A new method for the estimation of bearing health state and remaining useful life based on the moving average cross-correlation of power spectral density. *Mech. Syst. Signal Process.* **2020**, *139*, 106617. [CrossRef]
14. Qian, Y.; Yan, R.; Gao, R.X. A multi-time scale approach to remaining useful life prediction in rolling bearing. *Mech. Syst. Signal Process.* **2017**, *83*, 549–567. [CrossRef]
15. Ali, J.B.; Chebel-Morello, B.; Saidi, L.; Malinowski, S.; Fnaiech, F. Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. *Mech. Syst. Signal Process.* **2015**, *56*, 150–172.
16. Chelmiah, E.T.; McLoone, V.I.; Kavanagh, D.F. Wear State Estimation of Rolling Element Bearings using Support Vector Machines. In Proceedings of the 2020 15th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 6–9 December 2020; Volume 1, pp. 306–311.
17. Choudhary, A.; Goyal, D.; Shimi, S.L.; Akula, A. Condition monitoring and fault diagnosis of induction motors: A review. *Arch. Comput. Methods Eng.* **2019**, *26*, 1221–1238. [CrossRef]
18. Hashemian, H.M. State-of-the-Art Predictive Maintenance Techniques. *IEEE Trans. Instrum. Meas.* **2011**, *60*, 226–236. [CrossRef]
19. Martins, J.F.; Pires, V.F.; Pires, A. Unsupervised neural-network-based algorithm for an on-line diagnosis of three-phase induction motor stator fault. *IEEE Trans. Ind. Electron.* **2007**, *54*, 259–264. [CrossRef]
20. Elforjani, M.; Shanbr, S. Prognosis of bearing acoustic emission signals using supervised machine learning. *IEEE Trans. Ind. Electron.* **2017**, *65*, 5864–5871. [CrossRef]
21. Scanlon, P.; Kavanagh, D.F.; Boland, F.M. Residual Life Prediction of Rotating Machines Using Acoustic Noise Signals. *IEEE Trans. Instrum. Meas.* **2013**, *62*, 95–108. [CrossRef]
22. Kavanagh, D.F.; Scanlon, P.; Boland, F. Envelope analysis and data-driven approaches to acoustic feature extraction for predicting the remaining useful life of rotating machinery. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 1621–1624.
23. Oumaamar, M.E.K.; Maouche, Y.; Boucherma, M.; Khezzar, A. Static air-gap eccentricity fault diagnosis using rotor slot harmonics in line neutral voltage of three-phase squirrel cage induction motor. *Mech. Syst. Signal Process.* **2017**, *84*, 584–597. [CrossRef]
24. Jiménez Espadafor, F.J.; Becerra Villanueva, J.A.; Palomo Guerrero, D.; Torres García, M.; Carvajal Trujillo, E.; Fernández Vacas, F. Measurement and analysis of instantaneous torque and angular velocity variations of a low speed two stroke diesel engine. *Mech. Syst. Signal Process.* **2014**, *49*, 135–153. [CrossRef]
25. Zhang, R.; Gu, F.; Mansaf, H.; Wang, T.; Ball, A.D. Gear wear monitoring by modulation signal bispectrum based on motor current signal analysis. *Mech. Syst. Signal Process.* **2017**, *94*, 202–213. [CrossRef]
26. Gu, F.; Wang, T.; Alwodai, A.; Tian, X.; Shao, Y.; Ball, A. A new method of accurate broken rotor bar diagnosis based on modulation signal bispectrum analysis of motor current signals. *Mech. Syst. Signal Process.* **2015**, *50–51*, 400–413. [CrossRef]
27. Hong, W.; Li, T.; Wang, S.; Zhou, Z. A General Framework for Aliasing Corrections of Inductive Oil Debris Detection Based on Artificial Neural Networks. *IEEE Sens. J.* **2020**, *20*, 10724–10732. [CrossRef]
28. Loutas, T.; Roulias, D.; Pauly, E.; Kostopoulos, V. The combined use of vibration, acoustic emission and oil debris on-line monitoring towards a more effective condition monitoring of rotating machinery. *Mech. Syst. Signal Process.* **2011**, *25*, 1339–1352. [CrossRef]
29. Delgado-Prieto, M.; Carino-Corrales, J.A.; Saucedo-Dorantes, J.J.; de Jesus Romero-Troncoso, R.; Osornio-Rios, R.A. Thermography-Based Methodology for Multifault Diagnosis on Kinematic Chain. *IEEE Trans. Ind. Inform.* **2018**, *14*, 5553–5562. [CrossRef]

30. Bagavathiappan, S.; Lahiri, B.; Saravanan, T.; Philip, J.; Jayakumar, T. Infrared thermography for condition monitoring—A review. *Infrared Phys. Technol.* **2013**, *60*, 35–55. [CrossRef]
31. Combet, F.; Zimroz, R. A new method for the estimation of the instantaneous speed relative fluctuation in a vibration signal based on the short time scale transform. *Mech. Syst. Signal Process.* **2009**, *23*, 1382–1397. [CrossRef]
32. Prudhom, A.; Antonino-Daviu, J.; Razik, H.; Climente-Alarcon, V. Time-frequency vibration analysis for the detection of motor damages caused by bearing currents. *Mech. Syst. Signal Process.* **2017**, *84*, 747–762. [CrossRef]
33. Hebda-Sobkowicz, J.; Zimroz, R.; Pitera, M.; Wyłomańska, A. Informative frequency band selection in the presence of non-Gaussian noise—A novel approach based on the conditional variance statistic with application to bearing fault diagnosis. *Mech. Syst. Signal Process.* **2020**, *145*, 106971. [CrossRef]
34. Liu, H.; Li, L.; Ma, J. Rolling Bearing Fault Diagnosis Based on STFT-Deep Learning and Sound Signals. *Shock Vib.* **2016**, *2017*, 6127479. [CrossRef]
35. Gao, R.; Yan, R. Non-stationary signal processing for bearing health monitoring. *IJMR* **2006**, *1*, 18–40. [CrossRef]
36. Wang, D.; Zhao, X.; Kou, L.L.; Qin, Y.; Zhao, Y.; Tsui, K.L. A simple and fast guideline for generating enhanced/squared envelope spectra from spectral coherence for bearing fault diagnosis. *Mech. Syst. Signal Process.* **2019**, *122*, 754–768. [CrossRef]
37. Kang, M.; Kim, J.; Wills, L.M.; Kim, J. Time-Varying and Multiresolution Envelope Analysis and Discriminative Feature Analysis for Bearing Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2015**, *62*, 7749–7761. [CrossRef]
38. Borghesani, P.; Ricci, R.; Chatterton, S.; Pennacchi, P. A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions. *Mech. Syst. Signal Process.* **2013**, *38*, 23–35. [CrossRef]
39. Wang, X.; Tang, G.; Wang, T.; Zhang, X.; Peng, B.; Dou, L.; He, Y. Lkurtogram Guided Adaptive Empirical Wavelet Transform and Purified Instantaneous Energy Operation for Fault Diagnosis of Wind Turbine Bearing. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–19. [CrossRef]
40. Yoo, Y.; Baek, J.G. A Novel Image Feature for the Remaining Useful Lifetime Prediction of Bearings Based on Continuous Wavelet Transform and Convolutional Neural Network. *Appl. Sci.* **2018**, *8*, 1102. [CrossRef]
41. Wang, Z.; Zhang, Q.; Xiong, J.; Xiao, M.; Sun, G.; He, J. Fault Diagnosis of a Rolling Bearing Using Wavelet Packet Denoising and Random Forests. *IEEE Sens. J.* **2017**, *17*, 5581–5588. [CrossRef]
42. Qiu, H.; Lee, J.; Lin, J.; Yu, G. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *J. Sound Vib.* **2006**, *289*, 1066–1090. [CrossRef]
43. Ye, X.; Hu, Y.; Shen, J.; Chen, C.; Zhai, G. An Adaptive Optimized TVF-EMD Based on a Sparsity-Impact Measure Index for Bearing Incipient Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [CrossRef]
44. Chelmiah, E.T.; Kavanagh, D.F. Hilbert Marginal Spectrum for Failure Mode Diagnosis of Rotating Machines. In Proceedings of the IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society, Online, 13–16 October 2021; pp. 1–6.
45. Maurya, S.; Singh, V.; Verma, N.K. Condition Monitoring of Machines Using Fused Features from EMD-Based Local Energy with DNN. *IEEE Sens. J.* **2020**, *20*, 8316–8327. [CrossRef]
46. Lv, Y.; Yuan, R.; Song, G. Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing. *Mech. Syst. Signal Process.* **2016**, *81*, 219–234. [CrossRef]
47. Grasso, M.; Chatterton, S.; Pennacchi, P.; Colosimo, B.M. A data-driven method to enhance vibration signal decomposition for rolling bearing fault analysis. *Mech. Syst. Signal Process.* **2016**, *81*, 126–147. [CrossRef]
48. Ali, J.B.; Fnaiech, N.; Saidi, L.; Chebel-Morello, B.; Fnaiech, F. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* **2015**, *89*, 16–27.
49. Moshrefzadeh, A. Condition monitoring and intelligent diagnosis of rolling element bearings under constant/variable load and speed conditions. *Mech. Syst. Signal Process.* **2021**, *149*, 107153. [CrossRef]
50. Buzzoni, M.; D'Elia, G.; Mucchi, E.; Dalpiaz, G. A vibration-based method for contact pattern assessment in straight bevel gears. *Mech. Syst. Signal Process.* **2019**, *120*, 693–707. [CrossRef]
51. Wang, D. K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited. *Mech. Syst. Signal Process.* **2016**, *70–71*, 201–208. [CrossRef]
52. Liu, X.; Li, H. A New Method of Fault Diagnosis in Rolling Bearings. In Proceedings of the 2019 fourth International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Hohhot, China, 24–26 October 2019; pp. 120–1203.
53. Zhong, Y.X.; Fan, H.L.; Lu, J.P.; Pang, L.; Li, Y.F. Research on Fault Diagnosis of Rolling Bearing Based on Wavelet Packet Transform and IPSO-SVM. In Proceedings of the 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Bangkok, Thailand, 16–19 December 2018; pp. 1682–1686.
54. Zheng, J.; Pan, H.; Cheng, J. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. *Mech. Syst. Signal Process.* **2017**, *85*, 746–759. [CrossRef]
55. Soualhi, A.; Medjaher, K.; Zerhouni, N. Bearing Health Monitoring Based on Hilbert–Huang Transform, Support Vector Machine, and Regression. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 52–62. [CrossRef]
56. Li, X.; Elasha, F.; Shanbr, S.; Mba, D. Remaining Useful Life Prediction of Rolling Element Bearings Using Supervised Machine Learning. *Energies* **2019**, *12*, 2705. [CrossRef]
57. Wang, H.; Xu, J.; Yan, R.; Gao, R.X. A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 2377–2389. [CrossRef]

58. Guo, L.; Lei, Y.; Xing, S.; Yan, T.; Li, N. Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines With Unlabeled Data. *IEEE Trans. Ind. Electron.* **2019**, *66*, 7316–7325. [CrossRef]
59. Mao, W.; He, J.; Tang, J.; Li, Y. Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network. *Adv. Mech. Eng.* **2018**, *10*, 168781401881718. [CrossRef]
60. Lee, J.; Qiu, H.; Yu, G.; Lin, J.; Rexnord Technical Services. In *Bearing Data Set*; IMS, University of Cincinnati, NASA Ames Prognostics Data Repository: Mountain View, CA, USA, 2007.
61. Wang, B.; Lei, Y.; Li, N.; Li, N. A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings. *IEEE Trans. Reliab.* **2018**, *69*, 401–412. [CrossRef]
62. Nectoux, P.; Gouriveau, R.; Medjaher, K.; Ramasso, E.; Chebel-Morello, B.; Zerhouni, N.; Varnier, C. PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In Proceedings of the IEEE International Conference on Prognostics and Health Management, Detroit, MI, USA, 8–10 June 2020; pp. 1–8.
63. Lee, J.; Jiang, J. Time Series Multi-Channel Convolutional Neural Network for Bearing Remaining Useful Life Estimation. In Proceedings of the 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, 3–6 October 2019; pp. 408–410.
64. Khamoudj, C.E.; Tayeb, F.B.; Benatchba, K.; Benbouzid, M. Induction Machines Bearing Failures Detection and Diagnosis using Variable Neighborhood Search. In Proceedings of the IECON 2019–45th Annual Conference of the IEEE Industrial Electronics Society, Lisbon, Portugal, 14–17 October 2019; Volume 1, pp. 6010–6015.
65. Javed, K.; Gouriveau, R.; Zerhouni, N.; Nectoux, P. Enabling Health Monitoring Approach Based on Vibration Data for Accurate Prognostics. *IEEE Trans. Ind. Electron.* **2015**, *62*, 647–656. [CrossRef]
66. *ISO 1683:2015*; Acoustics—Preferred Reference Values for Acoustical and Vibratory Levels. International Organization for Standardization: Geneva, Switzerland, 2015.
67. Stanković, L.; Papandreou-Suppappola, A.; Iem, B.; Boudreaux-Bartels, G.; Katkovnik, V.; Djurović, I.; Matz, G.; Hlawatsch, F.; Napolitano, A. Chapter 9—Noise Analysis and Random Processes in the (t,f) Domain. In *Time-Frequency Signal Analysis and Processing*, 2nd ed.; Boashash, B., Ed.; Academic Press: Oxford, UK, 2016; pp. 521–573. [CrossRef]
68. Caesarendra, W.; Tjahjowidodo, T. A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing. *Machines* **2017**, *5*, 21. [CrossRef]
69. Wu, B.; Li, W.; Qiu, M.Q. Remaining Useful Life Prediction of Bearing with Vibration Signals Based on a Novel Indicator. *Shock Vib.* **2017**, *2017*, 8927937. [CrossRef]
70. Bellman, R. Dynamic Programming. *Science* **1966**, *153*, 34–37. [CrossRef]
71. Zhou, Z.; Mo, J.; Shi, Y. Data imputation and dimensionality reduction using deep learning in industrial data. In Proceedings of the 2017 third IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 2329–2333.
72. Bandyopadhyay, I.; Purkait, P.; Koley, C. Performance of a Classifier Based on Time-Domain Features for Incipient Fault Detection in Inverter Drives. *IEEE Trans. Ind. Informat.* **2019**, *15*, 3–14. [CrossRef]
73. Meng, L.; Miao, W.; Chunguang, W. Research on SVM Classification Performance in Rolling Bearing Diagnosis. In Proceedings of the 2010 International Conference on Intelligent Computation Technology and Automation, Changsha, China 11–12 May 2010; Volume 3, pp. 132–135.
74. Shi, R.; Ngan, K.N.; Li, S. Jaccard index compensation for object segmentation evaluation. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 4457–4461.
75. Niwattanakul, S.; Singthongchai, J.; Naenudorn, E.; Wanapu, S. Using of Jaccard coefficient for keywords similarity. In Proceedings of the International Multiconference of Engineers and Computer Scientists, Hong Kong, China, 13–15 March, 2013; Volume 1, pp. 380–384.
76. Deng, H.; Fannon, D.; Eckelman, M.J. Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata. *Energy Build.* **2018**, *163*, 34–43. [CrossRef]
77. Medar, R.; Rajpurohit, V.S.; Rashmi, B. Impact of Training and Testing Data Splits on Accuracy of Time Series Forecasting in Machine Learning. In Proceedings of the 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, India, 17–18 August 2017; pp. 1–6.
78. Sutrisno, E.; Oh, H.; Vasan, A.S.S.; Pecht, M. Estimation of remaining useful life of ball bearings using data driven methodologies. In Proceedings of the 2012 IEEE Conference on Prognostics and Health Management, Denver, CO, USA, 17–18 June 2012; pp. 1–7.
79. Singleton, R.K.; Strangas, E.G.; Aviyente, S. Extended Kalman filtering for remaining-useful-life estimation of bearings. *IEEE Trans. Ind. Electron.* **2014**, *62*, 1781–1790. [CrossRef]
80. Lei, Y.; Li, N.; Gontarz, S.; Lin, J.; Radkowski, S.; Dybala, J. A model-based method for remaining useful life prediction of machinery. *IEEE Trans. Reliab.* **2016**, *65*, 1314–1326. [CrossRef]