

An open set recognition methodology utilising discrepancy analysis for gear diagnostics under varying operating conditions

Stephan Schmidt^{a,*}, P. Stephan Heyns^a

^a*Centre for Asset Integrity Management (C-AIM), Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa*

Abstract

Historical fault data are often difficult and expensive to acquire, which can prohibit the application of supervised learning techniques in the condition-based maintenance field. Hence, novelty detection techniques such as discrepancy analysis are useful because only healthy historical data are required. However, even if discrepancy analysis is implemented on a machine, some historical fault data will become available during the operational lifetime of the machine and could be utilised to improve the efficiency of the condition inference process. An open set recognition methodology relying on discrepancy analysis is proposed that is capable of detecting novelties when only healthy historical data are available. It is also capable of inferring the condition of the machine if historical fault data are available and it is further able to detect novelties in regions not well supported by the historical fault data. In the condition recognition procedure, Gaussian mixture models are used with Bayes' rule and a decision rule to infer the condition of the machine in an open set recognition framework, where it is emphasised that it is beneficial to use the complete datasets (i.e. data acquired throughout the life of the component) for optimising the models. The benefit of the open set recognition model is that it is easy to incorporate new historical data in the framework as the data become available. In this work, practical aspects of the condition inference process such as the importance of good decision boundaries are highlighted and discussed as well. The methodology is validated on a synthetic dataset and experimental datasets acquired under varying operating conditions and it is also compared to a conventional classification process where discrepancy analysis is not used. The results indicate the potential of using the proposed methodology for rotating machine diagnostics under varying operating conditions.

Keywords:

*Corresponding author.

Email address: stephanschmidt@zoho.com (Stephan Schmidt)

1. Introduction

Reliable condition monitoring techniques are essential when performing condition-based maintenance on expensive rotating machine assets [1, 2]. Advanced signal processing [3–13] and sophisticated supervised machine learning techniques [14–23] are actively investigated to improve the condition monitoring task. Deep learning techniques have also recently been used to not only infer the condition of the machine, but also to extract features from the raw dataset i.e. it is not necessary to manually extract the features [24, 25].

However, many supervised machine learning techniques assume that all class labels are available at time of training and therefore the techniques require much historical fault data to be available, which can be expensive and impractical to acquire. This has resulted in novelty detection techniques [26–28] and techniques which address the class imbalance between healthy and damaged states [25] to be investigated for machine condition monitoring applications. Another shortcoming of standard supervised learning techniques in the condition monitoring field is that the condition classification task is implicitly addressed as a Closed Set Recognition (CSR) problem as opposed to an Open Set Recognition (OSR) problem. In the OSR framework, it is assumed that the historical datasets are samples from a population of damage modes and therefore the class label can only be assigned to data that are well supported by a class from the historical dataset [29–34]. This is in contrast to a CSR framework where it is assumed that the class label can only originate from the class labels in the historical dataset and is implicitly used in most supervised machine learning approaches for condition monitoring.

The differences between the decision boundaries of CSR and OSR frameworks are shown in Figure 1(a) and Figure 1(b) for a machine with healthy historical data and historical fault data of two damage modes. In the OSR framework (i.e. Figure 1(b)), predictions are only made in regions supported by historical data as opposed to the CSR (i.e. Figure 1(a)). Hence, the CSR framework can lead to erroneous results when used for data from a new class or data not supported

Abbreviations: CCP, Conventional Classification Procedure; CSR, Closed Set Recognition; CWT, Continuous Wavelet Transform; DR, Decision Rule; EM, Expectation Maximisation; GMM, Gaussian Mixture Model; GNB, Gaussian Naive Bayes; LDA, Linear Discriminant Analysis; LR, Logistic Regression; NLL, Negative Log-Likelihood; PCA, Principal Component Analysis; QDA, Quadratic Discriminant Analysis.

by the historical data, i.e. outliers. However, it has to be emphasised that both CSR and OSR

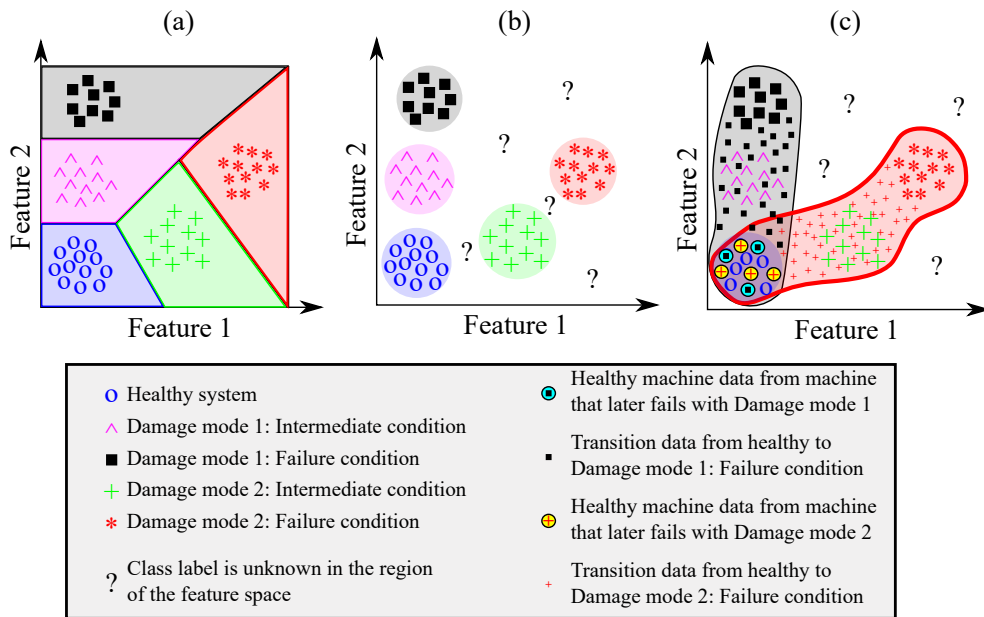


Figure 1: The differences between the decision boundaries of a Closed Set Recognition (CSR) framework and an Open Set Recognition (OSR) framework are shown in (a) and (b) respectively for artificial data. The proposed decision boundaries that are required for an ideal OSR framework for machine condition monitoring are shown in (c). In (c), the five classes used in (a) and (b) are reduced to three classes i.e. healthy, damage mode 1 and damage mode 2. The intermediate condition refers to a machine that is somewhere between a healthy and a failure state.

frameworks are supervised learning problems where the labels are used in the loss function during model optimisation. The key distinction between the aforementioned frameworks is the assumption that is made of the class labels; if a complete representation of the potential class labels is available at the time of training, it is a CSR problem, otherwise, it is a OSR problem. The open set recognition problem does not only compromise the ability of conventional classifiers to correctly predict the condition of the machine, it can also potentially compromise the effectiveness of automatic feature extraction methods. This is because the automatic feature extraction methods for example are optimised based on the data available at the time of training the model and may not be the optimal features to separate new or unseen damage modes in the feature space.

Throughout the operational lifetime of a machine, the machine transitions from a healthy state to some damaged state e.g. a crack initialises and grows. However, this results in problems when using the OSR framework shown in Figure 1(b); only the discrete conditions can be labelled with a class label, with the transitions between conditions labelled as novelties when the conditions are well-separated in the feature space. Hence, it is necessary to exploit the complete dataset (i.e. as the machine transitions from a healthy state to a damaged state) to learn the transition

path of the features so that the correct label can be assigned as the machine transitions between conditions. This results in decision regions shown in Figure 1(c). It is also easier to assign class labels to the approach in Figure 1(c) as opposed to 1(b), because the damage initiation time and the intermediate states may not be known for the historical fault data and can be difficult to estimate without stopping and disassembling the machine.

Discrepancy analysis is a powerful novelty detection technique that allows gearbox diagnostics to be performed under varying operating conditions [35–39]. In discrepancy analysis, a localised novelty score or discrepancy measure is assigned to the extracted localised features with a model optimised on the features of a healthy gearbox. The discrepancy measure calculated for the features at each angle index is concatenated to form a discrepancy signal, which can subsequently be processed to infer the condition of the component. Different inputs to the model e.g. windowed vibration data [36], features extracted from the wavelet packet transform [39], features extracted from the continuous wavelet transform [37, 38] as well as different models e.g. Gaussian mixture models [36], Gaussian models [39], hidden Markov models [38] can be used in the process. The synchronous average of the discrepancy signal is a very efficient technique for gear diagnostics under varying operating conditions [35–38]. Despite the fact that discrepancy analysis can be used to infer the condition of the machine and to characterise and trend damage, historical fault data will become available during the operational lifetime of rotating machines and need to be utilised to improve the condition monitoring task. Hence, the conventional discrepancy analysis approach to novelty detection has to be extended so that historical fault data can be incorporated into the condition inference process. This can be achieved by carefully designing a discrepancy analysis-based condition recognition methodology which is able to

- operate when only healthy data are available e.g. for new machines without historical fault information available,
- automatically infer the condition of the machine if historical fault data become available,
- include historical fault data from a new damage mode in the condition recognition procedure i.e. without re-optimising the model on all of the historical data, and
- label data from new damage modes or data that are not well supported by historical data as novelties i.e. it should be addressed as an OSR problem as opposed to a CSR problem.

In this work, a condition recognition methodology is proposed that satisfies the aforementioned properties by using discrepancy analysis as basis for automatic novelty detection and condition inference. The main contributions of this research are summarised as follows:

- An OSR methodology using discrepancy analysis is proposed that can perform automatic novelty detection and condition inference with probabilistic inference techniques.
- The methodology relies on discrepancy analysis which is capable of performing gearbox diagnostics under varying operating conditions [36–38]. The discrepancy signal is sensitive to damage under varying operating conditions because fault sensitive features are used with a model of the healthy features to generate the signal. Therefore, the features extracted from the discrepancy signal are potentially sensitive to damage under varying operating conditions as well. Therefore the methodology can be used for condition monitoring on machines operating under varying conditions.
- It is demonstrated that the classification task is simplified by using discrepancy analysis as a pre-processing step.
- It is demonstrated and emphasised that the classification accuracy does not necessarily convey the actual performance of the classifier and that the characteristics of the decision boundaries must be considered when dealing with condition monitoring problems.
- It is demonstrated that it is beneficial to use all of the available historical data as training data (i.e. from a healthy machine to a failure), as opposed to defining discrete states for each class or condition as shown in Figure 1(b), for an open set recognition problem. This results in the trajectory of the data to be learned as the condition of the machine changes from healthy to a specific damage mode as shown in Figure 1(c).

The proposed methodology is presented in Section 2, whereafter the condition recognition part of the methodology is investigated on synthetic data in Section 3. The methodology is validated on experimental data and compared to a conventional classification procedure that does not use discrepancy information in Section 4. Finally, the work is concluded and recommendations are made for future investigations in Section 5.

2. Methodology

The proposed methodology is presented in Figure 2. Vibration data and rotational speed in-

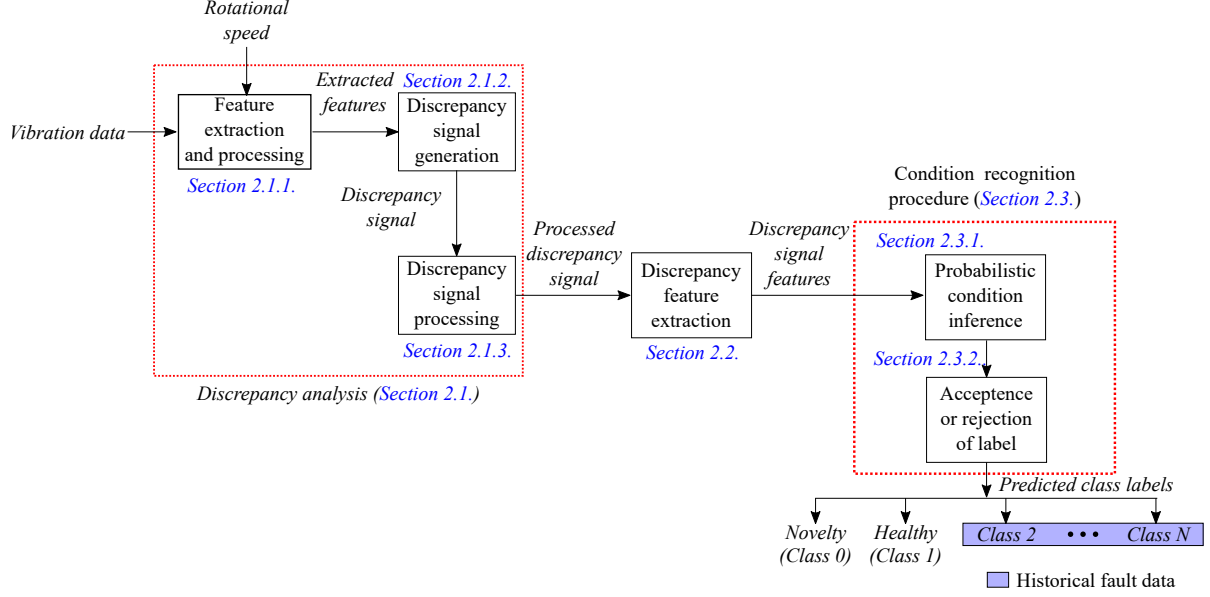


Figure 2: The proposed automatic novelty detection and condition inference methodology is presented, with the sections containing each part of the methodology given as well. In this figure, it is assumed that there are historical data of $N - 1$ damage modes available, with the class labels used consistently throughout this paper e.g. a class label of 0 will always indicate a novelty or a rejected class label.

formation of the machine component are converted to a processed discrepancy signal using the discrepancy analysis framework. The processed discrepancy signal is rich with diagnostic information and therefore discrepancy features are extracted from the processed signal. The extracted discrepancy signal features serve as an input to an OSR condition recognition system which is capable of performing novelty detection and inferring the condition of the machine under varying operating conditions. The rotational speed information can be acquired from tachometers or it can be estimated from the vibration data itself.

More information on each step in Figure 2 is given in the subsequent sections. The proposed methodology is a proof of concept and therefore different features and models can be used if desired.

2.1. Discrepancy analysis

Discrepancy analysis uses a model of the features, extracted from data that were acquired from a healthy machine, to assign discrepancy measures to the features of new data. This discrepancy measure is used to form a discrepancy signal which is analysed using techniques such as synchronous averaging to infer the condition of the gearbox [35–38]. A brief overview of the features and models used in this paper are given in this section.

2.1.1. Feature extraction for discrepancy analysis

In this work, the feature extraction procedure for the discrepancy analysis step, is very similar to the procedure used in Ref. [38]. In discrepancy analysis, it is necessary for the features to be described in terms of time or shaft angle and therefore time-frequency analyses methods are frequently employed [37–39]. Hence, the Continuous Wavelet Transform (CWT), which is popular for analysing non-stationary signals for rotating machine diagnostics [4, 12, 23, 40], is used with a Mexican hat basis function to obtain an angle-order distribution of the order tracked vibration signal, with the order tracking performed with the available rotational speed (or phase) information. The rotational speed can be estimated from measurement equipment [41] or estimated from the vibration signal [42, 43] depending on the available equipment. The wavelet coefficients of the first three gear mesh components are extracted for different shaft angles, whereafter a windowing procedure is used to extract localised features from the wavelet coefficient signals. The windowing procedure is described in detail in Refs. [38, 39]. In this work, ninety rectangular windows per gear revolution with an overlap of 50% between coinciding windows are used on the wavelet coefficients of each gear mesh component. The features extracted from each window are given in Table 1. The ten statistical

Table 1: The statistical features calculated for the windowed wavelet coefficients of each gear mesh frequency, with more details on the features found in Refs. [18, 19, 44].

1	Mean	2	Standard deviation
3	RMS	4	Maximum
5	Kurtosis	6	Skewness
7	Shape factor	8	Impulse factor
9	Crest factor	10	Clearance factor

features in Table 1, applied to the wavelet coefficients of the three gear mesh components, result in a 30 dimensional feature space and are denoted by $\mathbf{b}_i \in \mathbb{R}^{30}$ for the features extracted from window i . Changes in the localised statistics of the wavelet coefficients, such as the localised impulsiveness for example, as opposed to changes in the values of the wavelet coefficients are detected by the healthy feature model. The benefit of this is illustrated in the paper by Schmidt et al. [39], where the localised features can enhance the diagnostic information in the wavelet coefficients.

2.1.2. Discrepancy signal generation

The discrepancy signal is generated by firstly modelling the features extracted from the healthy data, whereafter a discrepancy measure is assigned to the new features to construct a discrepancy signal. A Gaussian Mixture Model (GMM) is used to model the features extracted with the process

in Section 2.1.1, because it is able to represent any distribution to an arbitrary level of accuracy if a sufficient number of mixture components is used [45] and it has performed well for discrepancy analysis [36]. Heyns et al. [36] did not find an improvement by incorporating the operating conditions in the GMM for discrepancy analysis and therefore only the fault features are used to optimise the GMM in this section.

The GMM parameters are obtained from the Expectation Maximisation (EM) algorithm, which is very efficient in iteratively obtaining the maximum likelihood estimates of the model parameters [45]. The Negative Log-Likelihood (NLL) of a GMM with N_c mixture components [45]

$$\eta[l] = -\log \left(\sum_{j=1}^{N_c} \pi_{\mathbf{b},j} \mathcal{N}(\mathbf{b}_l | \boldsymbol{\mu}_{\mathbf{b},j}, \boldsymbol{\Sigma}_{\mathbf{b},j}) \right), \quad (1)$$

is used as a discrepancy measure, where \mathbf{b}_l denotes the features extracted from window l , $\pi_{\mathbf{b},j}$ is the mixture coefficient, $\boldsymbol{\mu}_{\mathbf{b},j}$ is the mean and $\boldsymbol{\Sigma}_{\mathbf{b},j}$ is the covariance of mixture component j of the GMM. The mixture component $\pi_{\mathbf{b},j}$ weights the contribution of the j th Gaussian model to the probability density function of the features and can be interpreted as the prior probability that component j is active in the GMM [45]. The mean $\boldsymbol{\mu}_{\mathbf{b},j}$, covariance $\boldsymbol{\Sigma}_{\mathbf{b},j}$ and mixture coefficient $\pi_{\mathbf{b},j}$ of each component j are the unknown parameters of the GMM and are estimated with the EM algorithm on the healthy features. The number of latent states or mixture components, denoted by N_c , is obtained by performing cross-validation. A multivariate Gaussian distribution over \mathbf{b} is denoted by $\mathcal{N}(\mathbf{b} | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ in Equation (1), where its mean and covariance are denoted by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.

2.1.3. Discrepancy signal post-processing

The synchronous average is a useful tool for analysing whether the discrepancy signal contains deterministic components attributed to gear fault impacts [36]. The synchronous average of the discrepancy signal of measurement j , denoted by $\mathbf{m}_\eta^{(j)} \in \mathbb{R}^{N_s}$, is calculated with

$$m_\eta^{(j)}[k] = \frac{1}{N_r} \sum_{i=1}^{N_r} \eta[k + (i-1)N_s], \quad (2)$$

where $1 \leq k \leq N_s$, the discrepancy signal is denoted by η , the number of rotations that was completed by the shaft is denoted by N_r and the number of samples per shaft revolution is an integer and denoted by N_s . The synchronous average can also be useful for attenuating the amplitude modulation due to non-cyclic stationary loads [46], which means that the synchronous averaged

discrepancy signal can be even more robust to non-cyclic stationary operating conditions.

In the next section, the features extracted from the processed discrepancy signal and subsequently used for the condition recognition procedure, are presented and discussed.

2.2. Discrepancy feature extraction

Two features are extracted from the synchronous average of the discrepancy signal given by equation (2) to infer the condition of the gear. The first feature

$$r_{j1} = \frac{\max(\mathbf{m}_\eta^{(j)}) - \min(\mathbf{m}_\eta^{(j)})}{\sqrt{\frac{1}{N_s} \left(\sum_{i=1}^{N_s} m_\eta^{(j)}[i]^2 \right)}}, \quad (3)$$

makes it possible to discern between localised changes and uniform changes in the synchronous average of the discrepancy signal. This is an important generalisation of the previous work in discrepancy analysis which was focused on localised faults. The second feature

$$r_{j2} = \max(\mathbf{m}_\eta^{(j)}) - \min(\mathbf{m}_\eta^{(j)}), \quad (4)$$

is the range of the synchronous average of the discrepancy signal and is sensitive to localised changes in the discrepancy signal. The features extracted from the synchronous average of measurement j is denoted by a two-dimensional vector $\mathbf{r}_j = [r_{j1}, r_{j2}]^T$. The two features, extracted from Equation (3) and Equation (4), have a low dimensionality which means that neither feature subset selection nor dimensionality reduction techniques are required to circumvent the curse of dimensionality or to visualise the features. The low-dimensionality of the feature space helps with understanding the performance of different classifiers for machine condition monitoring which subsequently supports the conclusions that are drawn from the results in this paper. However, more optimised features can certainly be investigated to obtain classifiers that perform better.

The DC offset or mean value of the synchronous average of a healthy machine can be dependent on the discrepancy model which can adversely affect the features extracted from the discrepancy signal. Hence, the synchronous average of the first measurement of a dataset is subtracted from the entire dataset to obtain synchronous averages that are roughly model independent, with the new synchronous averages used in the feature extraction process. This is allowed, because it is assumed that the machine starts from an approximately healthy condition and deteriorates over time.

2.3. Condition recognition procedure

In this section, the condition recognition procedure that uses the discrepancy signal features in Section 2.2 to infer the condition of the gear is presented. In the condition recognition procedure, the objective is to recognise that the data is part of a specific class and to classify or label the data accordingly. However, if the condition cannot be recognised, it means that either a new damage mode is present or an outlier is detected and the data is labelled as a novelty. This is implemented with a OSR model which consists of two steps; probabilistic condition inference is performed to predict the class label of the machine, whereafter the class label is either accepted or rejected.

2.3.1. Probabilistic condition inference

It is desired to cast the condition inference problem in a OSR framework as opposed to the conventional CSR framework. Scheirer et al. [30] formalised the OSR problem in their work and contrasted it to the CSR framework used in many applications. Casimir et al. [31] investigated a methodology using a nearest neighbour classifier with two rejection options for induction motor diagnostics, while Lazzaretti [29] investigated an OSR methodology for power distribution events. The work in Refs. [29, 30, 32, 33] used support vector machines to address the OSR problem, where Lazzaretti et al. [29] found a support vector data description model performs the best of the considered OSR models for power distribution event detection.

In the machine condition monitoring field, the machine starts from an approximately healthy condition and deteriorates to a specific damage mode until the machine is stopped or the component fails. Instead of searching for the crack initiation time in the historical data and discretizing the data into different conditions to create different damage classes which can result in classification problems as discussed previously, it is desired to learn the trajectory of the data by using the complete dataset for the OSR model optimisation problem. If the trajectory of the data is learned, it can be used to infer the condition of the machine at arbitrary damage states as seen in Figure 1(c). This is essential for applying OSR in the condition monitoring field.

The problem with learning the trajectory of the data is the resulting overlap between the data of the healthy class and the data from the damaged modes as seen in Figure 1(c). This implicitly violates the objectives in the support vector optimisation task i.e. finding a hyperplane that maximises the distance between two classes, and due to the fact that the trajectory needs to be learned, it is concluded that support vector machines are not ideal for this task. GMMs have been

successfully used for fault classification [47] and can potentially learn the trajectory of the data through the feature space. However, density-based methods such as GMMs suffer from the curse of dimensionality which adversely influence their performance in OSR problems [29]. The discrepancy features' dimensionality in condition monitoring is significantly lower than the dimensionality of most OSR applications and alleviates the problems of GMMs in OSR problems. Hence, GMMs are used with Bayes' rule for the OSR task in this work.

The condition is inferred from the discrepancy signal features by selecting the class with the highest posterior probability i.e. $\arg \max_i (P(C_i|\mathbf{r}))$, with the posterior probability of class C_i being obtained with Bayes' rule

$$P(C_i|\mathbf{r}) = \frac{p(\mathbf{r}|C_i)P(C_i)}{p(\mathbf{r})}. \quad (5)$$

The likelihood function and prior probability of class i are denoted by $p(\mathbf{r}|C_i)$ and $P(C_i)$ respectively, with the unconditional density over the features denoted by $p(\mathbf{r})$. A unique GMM likelihood function is used for each class, with the GMM associated with class i denoted by

$$p(\mathbf{r}|C_i) = \sum_{j=1}^{N_{c,r}} \pi_{\mathbf{r},j}^{(i)} \mathcal{N}(\mathbf{r}|\boldsymbol{\mu}_{\mathbf{r},j}^{(i)}, \boldsymbol{\Sigma}_{\mathbf{r},j}^{(i)}). \quad (6)$$

In Equation (6), the parameters of the GMM are estimated with the EM algorithm separately for each class i , with $\pi_{\mathbf{r},j}^{(i)}$, $\boldsymbol{\mu}_{\mathbf{r},j}^{(i)}$ and $\boldsymbol{\Sigma}_{\mathbf{r},j}^{(i)}$ denoting the parameters of the j th component of class i . The appropriate number of mixture components $N_{c,r}$ in the GMM for the condition recognition procedure is estimated by cross-validation. The complete dataset, i.e. from a healthy machine to a damaged machine, is used in this methodology to optimise the GMM with each damage mode, with only the healthy data being used to optimise the GMM of the healthy machine.

In Refs. [20, 21], the class with the largest likelihood for the dataset is the inferred class i.e. $\arg \max_i p(\mathbf{r}|C_i)$, which is the same as $\arg \max_i P(C_i|\mathbf{r})$ when each class has the same prior probability. The same prior probability cannot be assigned to each class in this framework, because there is a large overlap in the healthy data space when using the complete datasets of the damage modes for trajectory learning, which can lead to significant classification errors in the healthy data region. This problem can be resolved by giving the model that correspond to the healthy machine features preference by ensuring that its prior probability is larger than the other models i.e. $P(C_{\text{healthy}}) > P(C_i)$ for all $i \neq \text{healthy}$. This comes from the expectation that the machine is healthy, unless there is strong evidence that the machine is in a damaged condition. In this work

$P(C_i) = c$ for all $i \neq \text{healthy}$ and $P(C_{\text{healthy}}) = 2 \cdot c$, with the unknown constant c obtained by ensuring the prior is a valid probability distribution i.e. by solving $\sum_i P(C_i) = 1$.

The benefit of using Bayes' rule with the GMM is that it is efficient to incorporate historical data from a new damage mode into this framework. This is performed by optimising a GMM on the newly acquired data, whereafter the prior probabilities of all the models have to be recalculated with the prior probability of the new class i set to $P(C_i) = c$ as well. Thereafter, Equation (5) can be used to predict the class label.

2.3.2. Acceptance or rejection of inferred class label

The posterior prediction obtained with Bayes' rule in Equation (5) is dependent on the likelihood value of the other classes due to the unconditional density, and it does not account for other classes that can be present in the OSR problem [34]. Hence, a Decision Rule (DR) is used on the likelihood function of the model or class i to determine whether the model is capable of predicting the class label $s_i \geq 0$ or not $s_i < 0$, with

$$s_i = \log p(\mathbf{r}|C_i) - \alpha_i^{\text{DR}}. \quad (7)$$

The threshold and model fitness metric associated with the log-likelihood of class i is denoted by α_i^{DR} and s_i , respectively. If $s_i < 0$ for all i , it means that none of the models are capable of inferring the correct class and therefore the class label predicted in Section 2.3.1 is rejected and a novelty is detected. The posterior probabilities $P(C_i|\mathbf{r})$ corresponding to the detected novelties are invalid and are not used further in the analysis. If a novelty is detected by the condition recognition system, it does not necessarily indicate that a new damage mode is present, but the data should rather be investigated further by an expert in the field.

There are many criteria to select the threshold of class i , which will depend on the application e.g. a tighter threshold can be used for critical machinery. In this work, α_i^{DR} is selected so that 1% of the validation set is labelled as a novelty.

2.4. Final remarks on the methodology

If discrete conditions are used to define damage modes, performance metrics such as the class averaged accuracy and Youden's index can be used to quantify the performance of the OSR model [33]. However, in the methodology, the complete dataset is used for training the condition recognition system which poses some difficulties in quantifying the performance of the condition recognition system i.e. the damaged class contains healthy data and the condition varies significantly throughout

the dataset. Hence, the following procedure is used for evaluating the performance of the proposed methodology:

- The measurement number at which the damage is detected and the characteristics of the posterior probabilities of the classes are investigated.
- The characteristics of the decision boundaries of different classifiers are investigated and compared, which is possible due to the two-dimensional discrepancy signal feature space. This makes it possible to investigate the implications of different decision boundaries for detecting a deteriorating gear condition which helps to evaluate the performance of the associated classification models.
- The OSR abilities of the methodology is investigated by visualising the two-dimensional discrepancy signal feature space. This makes it possible to evaluate the possibility of the OSR to generalise for other damage modes.

3. Synthetic data investigation

An investigation is performed in this section on the discrepancy feature extraction and condition recognition parts of the methodology presented in Figure 2, where it is assumed that the synchronous average of the different damage modes are already available. The purpose of this investigation is

- to investigate the characteristics of the discrepancy features,
- to highlight the importance of the decision boundaries for machine condition monitoring,
- to highlight that it is beneficial to use the complete dataset as the gear transitions from a healthy to a damaged state in OSR condition monitoring problems,
- and to compare the performance of the OSR condition recognition system to other conventional CSR classifiers on the same data.

3.1. Synthetic dataset

For the synthetic dataset, synchronous averages of six classes are generated with the following equations

$$m1_{\eta}^{(j)}[l] = 0.1 \cdot \epsilon[l], \quad (8)$$

$$m2_{\eta}^{(j)}[l] = \kappa \cdot \beta \cdot \delta(l - \lfloor l_{med} \rfloor) + 0.1 \cdot \epsilon[l], \quad (9)$$

$$m3_{\eta}^{(j)}[l] = \kappa \cdot \beta + 0.1 \cdot \epsilon[l], \quad (10)$$

$$m4_{\eta}^{(j)}[l] = \kappa \cdot \beta \cdot \left(\delta(l - \lfloor l_{med} \rfloor) + \sum_{j=1}^3 \sum_{k=1}^2 (0.8 - 0.2 \cdot j) \cdot \delta \left(l - \lfloor l_{med} \rfloor - j \cdot (-1)^k \right) \right) + 0.1 \cdot \epsilon[l], \quad (11)$$

$$m5_{\eta}^{(j)}[l] = \kappa \cdot \beta \cdot \sum_{j=0}^{\lfloor N_s/3 \rfloor} \delta(j - l) + 0.1 \cdot \epsilon[l], \quad (12)$$

$$m6_{\eta}^{(j)}[l] = \kappa \cdot \beta + 0.1 \cdot \sin(10\pi \cdot l/N_s) + 0.1 \cdot \epsilon[l], \quad (13)$$

where for example, $m1$ denotes the synchronous average of Class 1 in Equation (8). In Equations (8)-(13), j denotes the measurement number, l denotes the position on the gear with $1 \leq l \leq N_s$ and $l \in \mathbb{Z}$. The median of l is denoted by l_{med} . The condition of the component is indicated with a scalar factor $0 \leq \kappa \leq 1$, where 0 indicates a healthy condition and 1 indicates the component is in its failure state according to the respective damage mode or class. The magnitude of the damaged component in its failure state, denoted β , is sampled from a uniform distribution \mathcal{U} with a domain of $[0.75, 0.90]$, i.e. $\beta \sim \mathcal{U}[0.75, 0.90]$. Zero mean Gaussian noise with a standard deviation of 1 is denoted by $\epsilon[l]$. The function $\delta(x) = 1$ for $x = 0$ and $\delta(x) = 0$ if $x \neq 0$ and the number of samples in the synchronous average is $N_s = 100$. The mean and standard deviation of the six synthetic synchronous average profiles are presented in Figure 3. Class 1 represents a healthy system, Class 2 and Class 4 represent localised damage such as a damaged gear tooth with slightly different characteristics [36, 38], Class 3 and Class 6 represent damage over the whole profile of the gear and lastly, Class 5 contains a relatively small portion of the gear that is damaged. The similarity between Class 2 and 4 and Class 3 and 6 is purposefully used to make it a challenging problem for the classifiers which will result in interesting decision regions.

The discrepancy features of the different classes, extracted with the procedure presented in Section 2.2 are presented in Figure 4(i) for the data in Figure 3. The assumption is made that the

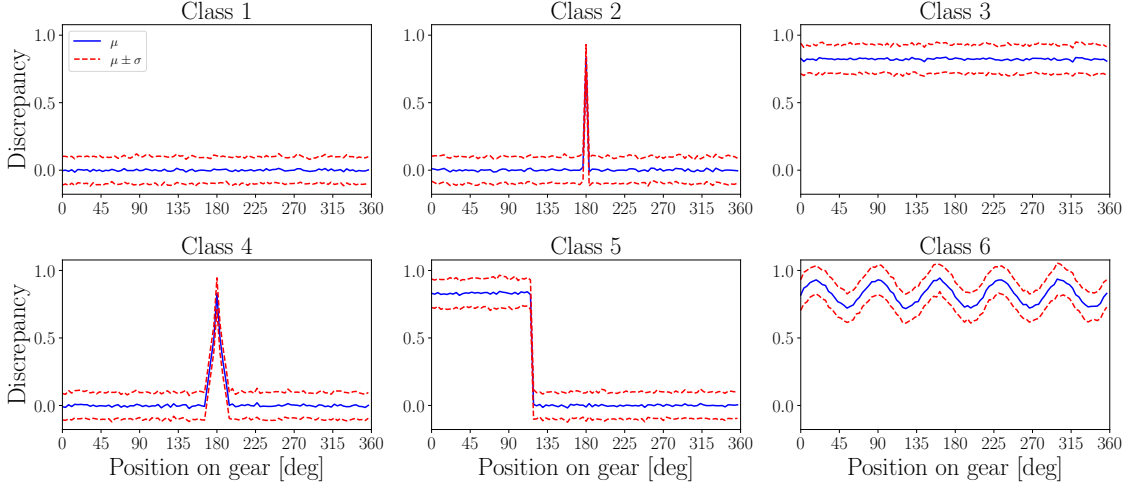


Figure 3: The mean μ and the standard deviation σ of 100 synchronous averages of the discrepancy signals of the six synthetic classes, investigated in Section 3. The data are generated with Equation (8) - (13).

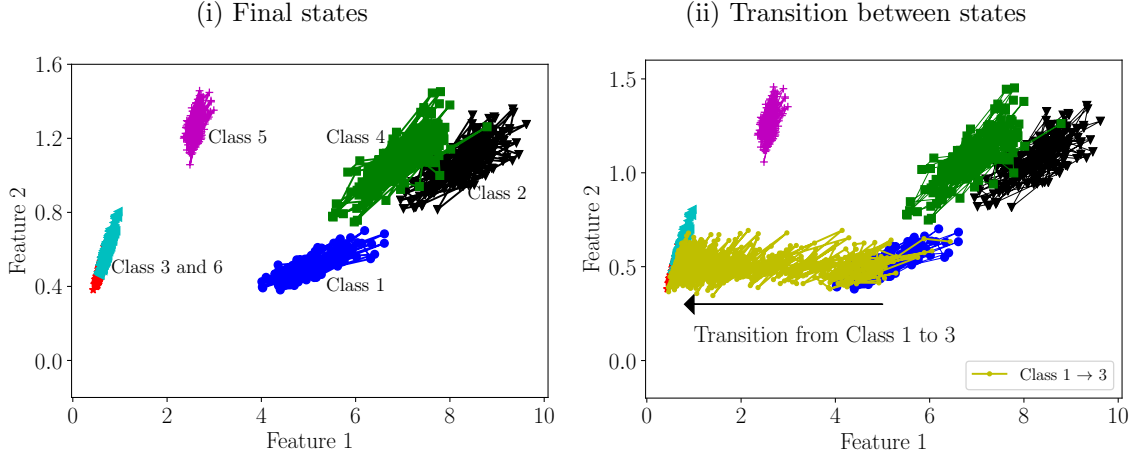


Figure 4: The features for the six classes in Figure 3 are presented in Figure 4(i). In Figure 4(ii), a stochastic transition from Class 1 to Class 3 is presented as an example.

system starts from a healthy condition (i.e. Class 1 in Figure 3 and Figure 4(i)) and stochastically transitions to the other conditions with an example shown in Figure 4(ii). The transition between Class 1 and Class i is performed by linearly increasing the factor κ in the appropriate Equation e.g. the κ in Equation (13) is changed for a transition between Class 1 and Class 6. The stochastic transition seen in Figure 4(ii) is because β in Equation (10) is a random variable.

In the next section, CSR classification models are compared to the OSR model with the data of the discrete classes in Figure 4(i) being used as training data.

3.2. Discrete machine conditions for model optimisation

In this section, a conventional classification process is used where discrete states (e.g. a gear with a broken tooth, a healthy gear etc.) are used as classes in the classification task to highlight the limitations of CSR models and the limitations of using discrete states for OSR models.

Due to the characteristics of the features in Figure 4(i), it is expected that Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Logistic Regression (LR) and Gaussian Naive Bayes (GNB) models will perform very well on the data and therefore the additional flexibility that neural networks and support vector machines provide are not required. The aforementioned models are implemented with scikit-learn [48] on the data in Figure 4(i) and compared to the OSR model discussed in Section 2.3.1 and Section 2.3.2. The training data and a query point are superimposed on the decision boundaries in Figure 5 for the LDA, QDA, LR and the OSR model. The query point highlights that the CSR classifiers predict a class label for data far from the training data and that the predictions are different due to the different characteristics of the decision boundaries. The classification error of the classifiers are given in Table 2 on the testing data, with the result of a GNB classifier presented as well. The LDA, QDA, LR and the OSR GMM classifiers perform equally well, with the GNB performing slightly worse. The classification error is attributed to the large overlap between Class 3 and Class 6. Even though the classification error is similar for

Table 2: Average classification error of the different models on the 200 testing measurements of each class, with the testing data generated similarly to the data in Figure 4(i).

LDA	QDA	GNB	LR	GMM
9.25%	9.92%	12.33%	9.25%	9.25%

the aforementioned models, their decision boundaries in Figure 5 have very different characteristics. The implication of this is that different predictions are made by the different classifiers as the data transition from Class 1 to another class. This emphasises that the classification errors are not the only important performance metric, but also the decision boundaries of the classifiers. The decision boundaries in a CSR framework are obtained from an optimisation process aiming to minimise the prediction error of the classifier for the historical data and may not necessarily have the desired properties for machine condition monitoring problems.

The superiority of the OSR GMM with three mixture components is seen in Figure 5(iv). The OSR GMM contains tightly fit boundaries around the regions supported by the historical data and a white region between the clusters which has a class label of 0. This indicates that the model

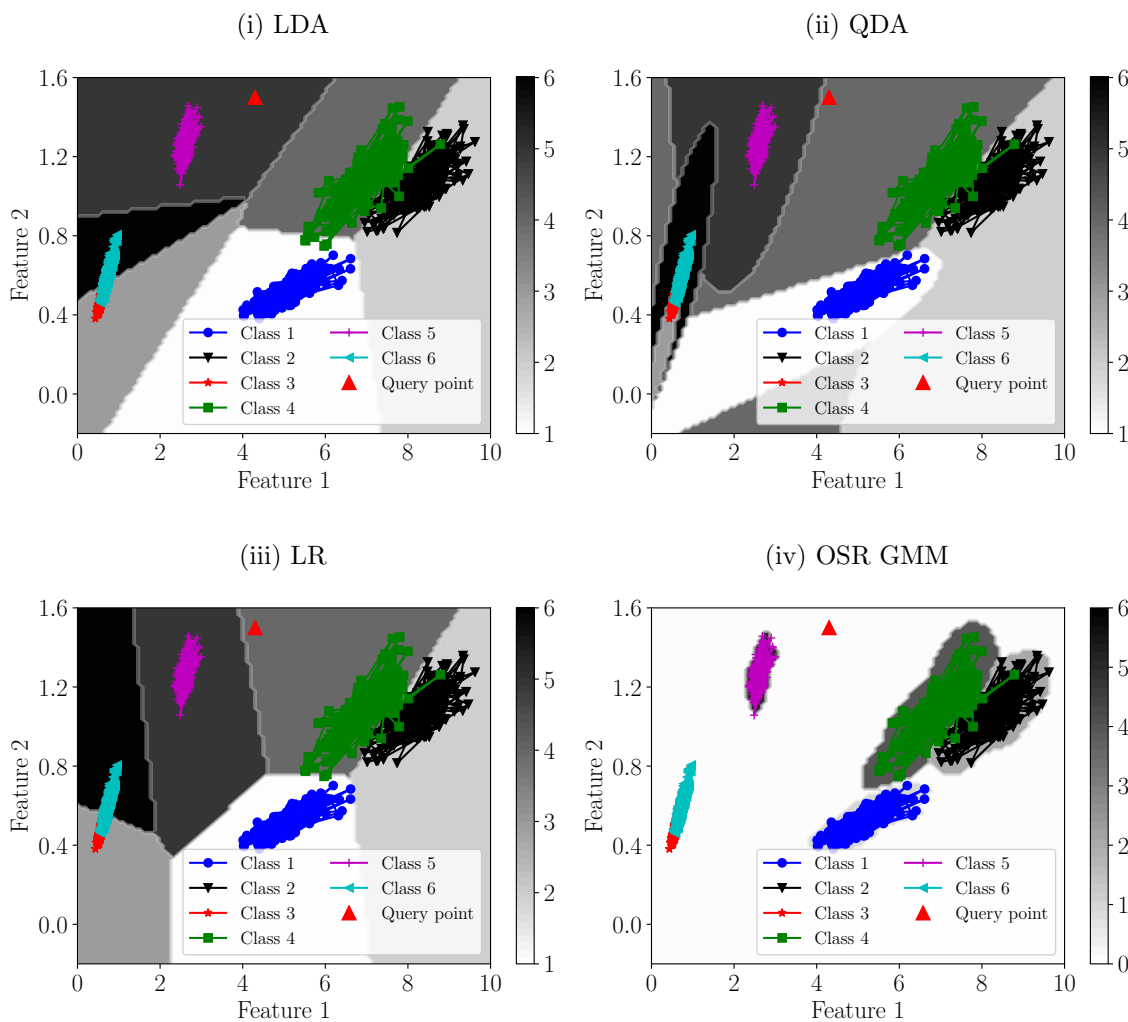


Figure 5: The decision regions for three classifiers using a CSR framework and the OSR model are superimposed on the training data in this figure. The colourbar indicates the class label, where a class label of 0 indicates that a novelty is detected in that region. Abbreviations: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Logistic Regression (LR) and Open Set Recognition Gaussian Mixture Model (OSR GMM).

is incapable of predicting the class label of the data in the regions that are not supported by the historical data and therefore novelties are detected in those regions. This is in contrast to the CSR classifiers, which can have misleading predictions in regions unsupported by the data.

The large novelty detection regions in Figure 5(iv) can however lead to problems when inferring the condition of a deteriorating machine, i.e. novelties will be detected in the transition process. A solution for this problem is therefore investigated in the next section.

3.3. Complete dataset for model optimisation

The OSR GMM decision boundaries in Figure 5(iv) are impractical for condition monitoring purposes because novelties will be detected almost continuously during the condition transition

process. Therefore, it is proposed in this article that the data of the transition from Class 1 to the other classes are used for optimising the GMM classifier i.e. the whole dataset, as it transitions from Class 1 to Class 3 in Figure 4(ii), is labelled as Class 3 in this section. Therefore, no distinctions are made between the initial state and the final state when optimising the model.

The number of mixture components in the GMM of each class was determined with five-fold cross-validation, with the number of components determined to be four. The training data and the resulting decision regions of the OSR GMM are presented in Figure 6. It can be observed that the

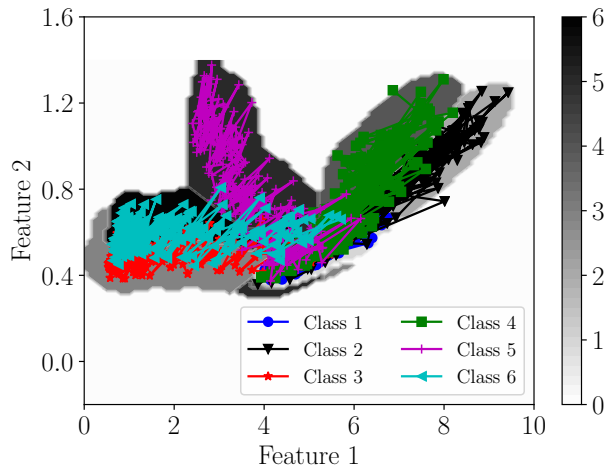


Figure 6: The decision regions for the OSR GMM that is optimised with the transition data is presented. For example, the data of the transition from Class 1 to Class 3 in Figure 4(ii) are the training data of Class 3 for the OSR GMM.

decision regions surround the training data, while the regions unsupported by the training data are classified as novelties i.e. a class label of 0 is obtained.

The posterior probability of the OSR GMM classifier using the complete dataset is compared to the posterior probability of the CSR LR classifier for the transition from Class 1 to Class 3 in Figure 7, with $\arg \max(P(C_i|\mathbf{r}))$ being used to infer the class label. A moving average with a length of 30 and an overlap of 25 measurements was used on the raw probability data to reduce the noise due to the stochastic transition process. The corresponding decision boundaries of the OSR GMM and CSR LR classifiers are shown in Figure 6 and Figure 5(iii) respectively. Initially, the data are in the healthy region and therefore the predicted class label is Class 1 for both classifiers. In Figure 7(i), the LR classifier incorrectly infers that Class 5 is present between measurement number 150 and 300 due to the characteristics of its decision boundary. In contrast to the LR classifier, the GMM condition recognition system in Figure 7(ii) correctly infers the class during the transition process. It is therefore clear from the results in Figure 6 and Figure 7(ii) that the decision regions

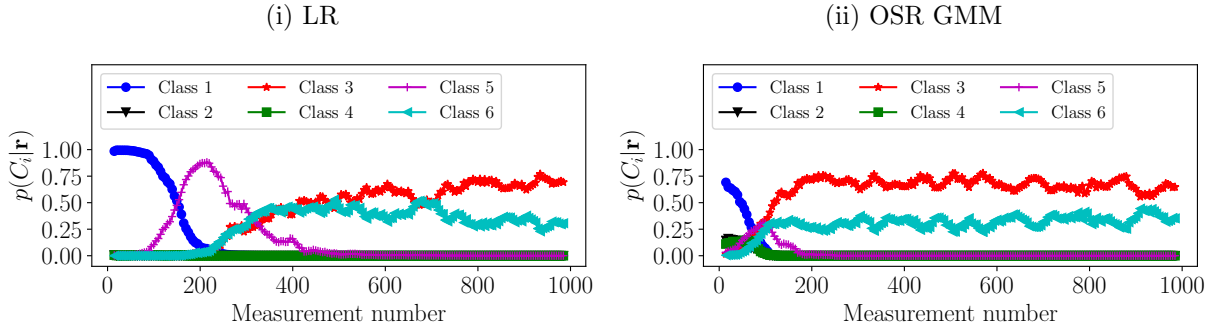


Figure 7: The posterior probability of the different classes for the LR classifier and the condition recognition system using the GMMs as the data stochastically transitions from Class 1 to Class 3. This is presented for the transition data in Figure 4(ii), using the decision boundaries in Figure 6 and Figure 5(iii).

can be controlled by learning the trajectory of the data with an appropriate classifier and the full training set (i.e. using continuous as opposed to discrete machine states), which is beneficial for the condition recognition process. This also emphasises that care should be taken when using non-linear classifiers that either transform the data to a high dimensional space e.g. support vector machines, or perform automatic feature selection e.g. neural networks. Using the aforementioned classifiers can result in the classes to be perfectly separated, but the decision regions may not preserve the important characteristics for machine condition monitoring applications.

The posterior probability in Figure 7(ii) does not indicate a change in condition between measurement number 200 and 1000 and can therefore not be used for fault trending. This is because the only aim of the condition recognition system is to determine whether the data belong to a specific class or not (when a novelty is detected). Therefore, it is suggested that the latent variables or hidden states of the GMM be used to detect changes in the features. The hidden states for the testing data, obtained from maximising the posterior distribution over the latent states as described in Ref. [45], are presented in Figure 8 for the model of Class 3, with three progression stages identified. The rapid alternation between two states indicate that the data reside in a region of the

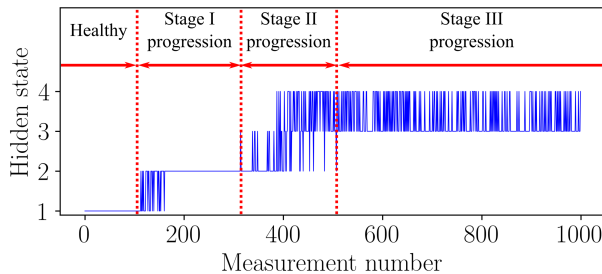


Figure 8: The hidden states of the GMM associated with the data of Class 3 as the data transition from Class 1 to Class 3.

feature space well supported by two clusters i.e. there is a large overlap between the two clusters. The first hidden state is associated with the healthy data region, whereafter the active hidden state change as the data move through the different clusters, which can support maintenance decisions. For example, the first progression stage can signal a warning and if the latent state of the GMM reaches the final progression stage i.e. stage 3, the machine needs to be stopped immediately.

The results in this section validate the abilities of the proposed discrepancy features, they highlight the benefits of using an OSR as opposed to a CSR framework, and the importance of the decision boundaries in the machine condition monitoring field is emphasised as well. In the next section, the methodology is validated on experimental data.

4. Experimental validation

The proposed methodology is validated on experimental gearbox data, acquired under varying operating conditions, in this section. An overview of the dataset is given in Section 4.1, whereafter the proposed methodology is investigated in Section 4.2 and compared to a conventional classification approach in Section 4.3.

4.1. Experimental dataset

The experimental setup, shown in Figure 9, consists of an alternator and an electrical motor which are independently controlled to exert specific operating conditions on the three helical gearboxes seen in the figure. Two healthy and two accelerated life tests were performed on the setup by

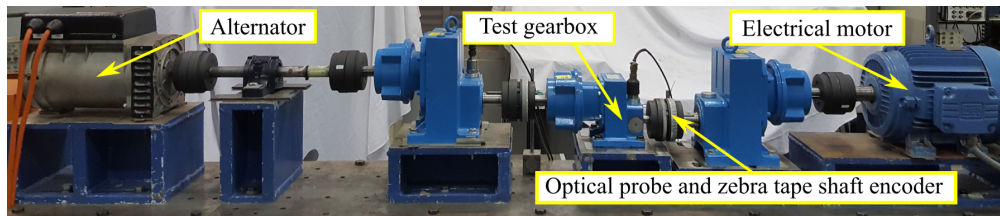


Figure 9: Experimental setup

introducing artificial damage on the gear of the test gearbox. The centre helical gearbox, labelled as the test gearbox in Figure 9, was monitored for damage throughout the fatigue tests by taking regular vibration measurements. The axial component of the tri-axial accelerometer, which is located on the bearing housing of the test gearbox, is used as source of vibration data. The optical probe and the zebra-tape shaft encoder are used to obtain the rotational speed information of the input shaft of the test gearbox and is used in the discrepancy analysis step of the methodology

shown in Figure 2. The operating conditions that were imposed on the system are presented in Figure 10. The duty cycle of the operating conditions is 10 seconds, with each measurement having

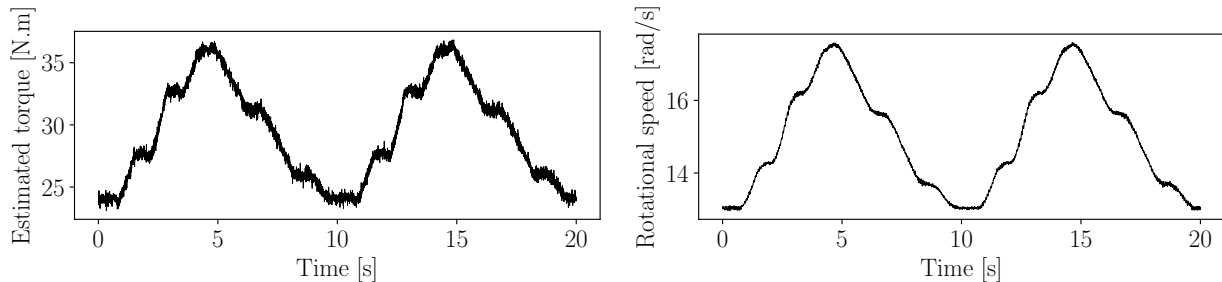


Figure 10: The operating conditions that were present for each measurement. The rotational speed is calculated from the geometrically compensated [41] tachometer signal.

a duration of 20 seconds. The sampling frequency of the vibration signals was 25.6 kHz and the sampling frequency of the tachometer signal was 51.2 kHz.

In the first experiment, vibration measurements were taken from a healthy test gearbox, whereafter the gearbox was disassembled and the gear was seeded with damage as shown in Figure 11(i). The gearbox was reassembled with the damaged gear and operated again with the operating conditions shown in Figure 10 until the gear tooth failed as shown in Figure 11(ii). In the second experiment, data were collected from a healthy gearbox whereafter the gearbox was disassembled and reassembled with a gear shown in Figure 11(iii). The gearbox was operated again, with the operating conditions in Figure 10 imposed on the system, until the damaged tooth ultimately failed as shown in Figure 11(iv). More severe damage was seeded in the gear in Figure 11(iii) compared to the gear in Figure 11(i) by creating a slot that was 20% deeper in the gear tooth. This resulted in a significantly shorter time of operation (i.e. 3 days as opposed to 20 days) and a more abrupt failure for the damaged gear in the second experiment.

In all subsequent sections where the data are shown over measurement number for the damaged gears, the first measurement was acquired for the gear in the initial condition (i.e. Figure 11(i) for experiment 1 or Figure 11(iii) for experiment 2) and the last measurement was acquired for the gears in their final condition (i.e. Figure 11(ii) for experiment 1 or Figure 11(iv) for experiment 2). In contrast, the healthy gears were in the same condition for all healthy gear measurements of the two experiments. The healthy gear data will be referred to as healthy or a class label of 1 will be given, while the damaged gears will be referred to as damaged or a class label of 2 will be given where necessary. This is to be consistent with the class labels in Figure 2. If more damage

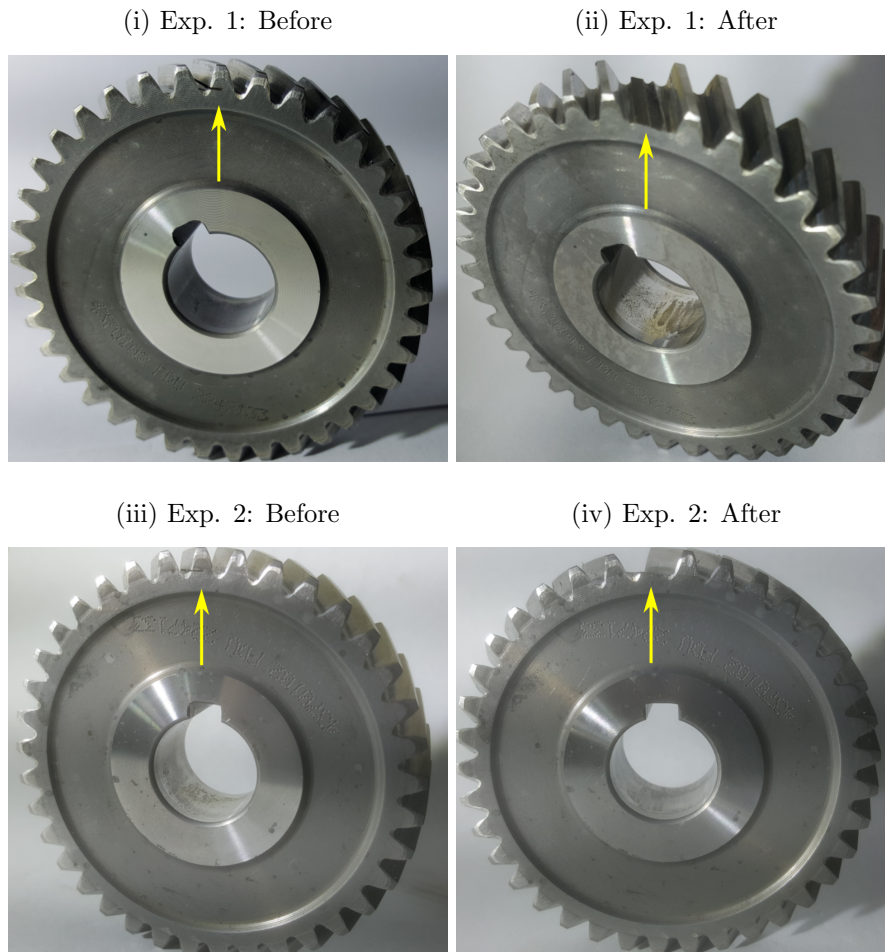


Figure 11: Damaged gears before and after the two fatigue experiments are presented with an arrow that indicates the location of the damage. The gear of the first experiment was tested for approximately 20 days before failure, in contrast to the second gear that ran approximately for 3 days before failing.

modes were to be investigated, it would be necessary to use more descriptive labels for the damaged datasets.

4.2. Application of proposed methodology

In this section, the proposed methodology in Section 2 is investigated on the experimental data of the Section 4.1. The first part of the investigation focuses on a novelty detection problem where only healthy historical data are available and the second part focuses on a novelty detection and classification problem where there are historical fault data available as well. A summary of the data used in the investigations of this section is presented in Table 3. Discrepancy analysis, the first step of the methodology in Figure 2, is investigated for the experimental datasets in the next section.

Table 3: The number of measurements used for the different datasets in the investigations, with the training and testing datasets being unique. The novelty detection training data refer to the case where the condition recognition system is only optimised with healthy data, where the classification training data refer to the case where the condition recognition system is optimised with historical fault data as well.

Dataset description	Training (Novelty detection)	Training (Classification)	Testing (Novelty detection & Classification)
Exp. 1: Healthy	44	44	50
Exp. 2: Healthy	0	0	50
Exp. 1: Damaged	0	166*	500*
Exp. 2: Damaged	0	0	215*

*The measurements are evenly spaced throughout the life of the gear.

4.2.1. Discrepancy analysis

The discrepancy analysis step is implemented using the procedure described in Section 2.1 on the healthy historical experimental data. The number of latent states in the GMM is estimated by using five-fold cross validation which resulted in eight latent states i.e. $N_c = 8$ in Equation (1) to be used. The synchronous average of a healthy gearbox and a gearbox with localised gear damage, calculated with Equation (2), are presented in Figure 12. The localised damage at 180° is easily

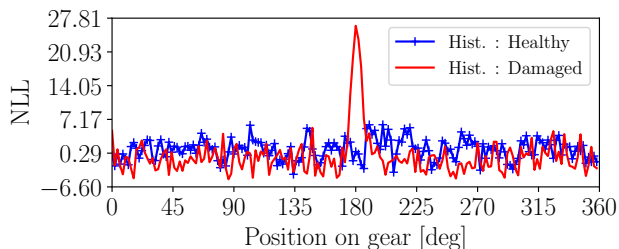


Figure 12: The synchronous average of the discrepancy signals of a healthy gearbox and of a gearbox with a damaged tooth at 180° is presented.

identified when compared to the synchronous average of a healthy gearbox and looks similar to Class 2 and Class 4 in Figure 3.

Heyns et al. [36] illustrated that discrepancy analysis is better suited than using the synchronous average or power spectral density of the vibration signal for gearbox diagnostics under varying operating conditions. This is illustrated by comparing the synchronous averaged vibration signal and the synchronous averaged discrepancy signal in Figure 13 for two measurements of the damaged gearbox of the first experiment. It is evident that the synchronous averaged discrepancy signal is more sensitive to damage than the synchronous averaged vibration signal. Therefore, it is sensible to use the synchronous averaged discrepancy signal as opposed to the synchronous averaged vibration signal in the next step.

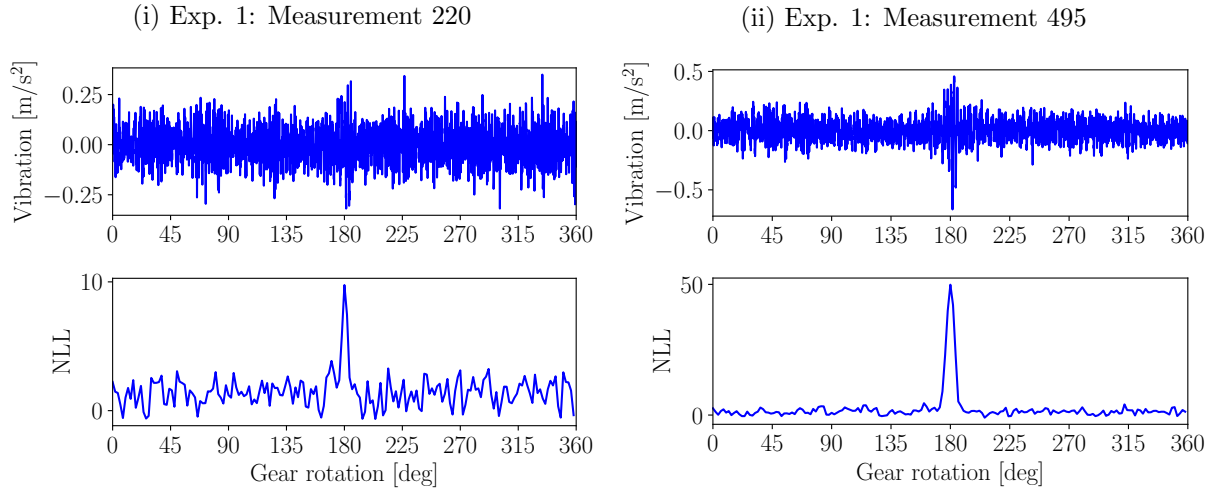


Figure 13: Synchronous averaged vibration signals are compared to synchronous averaged discrepancy signals of a gearbox with a damaged tooth at 180° for two measurements.

4.2.2. Discrepancy feature space of experimental dataset

The discrepancy features, discussed in Section 2.2, are presented in Figure 14 for the four gear datasets presented in Section 4.2 and Table 3. In Figure 14(i), all the datasets start at the zoomed portion box, also indicated by "Initial" and as the gear condition changes, the features move to "Final" with a trajectory in a similar direction as Class 2 and Class 4 in Figure 4(i). The large

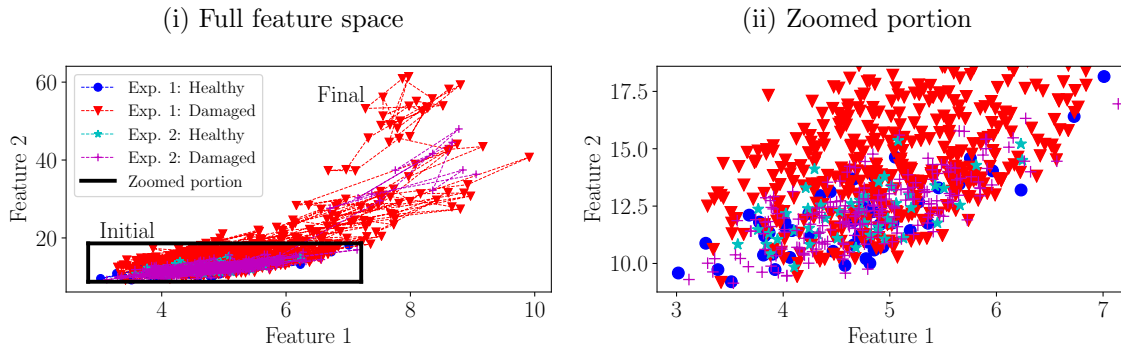


Figure 14: The full feature space of the four experiments is presented, with a zoomed view of the healthy feature region as well. In Figure 14(i), the starting position of the transition process, denoted "Initial", and the stopping position of the transition process, denoted "Final", are presented for clarity.

overlap in the healthy data region in Figure 14(ii) is attributed to the fact that the damaged gears start approximately in a healthy condition, i.e. the seeded damage is small, whereafter the gears deteriorate during the experiments. The influence of the deterioration of the gear is clearly seen in Figure 14(i), with the two gears having approximately the same trajectory. The slight differences in the trajectories are attributed to the fact that the seeded damage had different

characteristics which resulted in a more abrupt failure for the second gear. The disassembling and reassembling process between the experiments also resulted in the features to have slightly different characteristics, however, the effect of this is negligible in comparison to the changes in the features as the condition of the gear changes.

In the next section, the condition recognition system is only optimised on healthy data, to test its ability to perform novelty detection.

4.2.3. Classification with only healthy historical data (Novelty detection)

It is necessary to test the ability of the condition recognition system to perform novelty detection when only historical data from a healthy machine are available. Therefore, it is assumed that only the healthy historical data in Table 3 are available for optimising the classifier presented in Figure 2. The discrepancy features are extracted from the aforementioned healthy historical data with the procedure discussed in Section 2.2, whereafter a single GMM is optimised on the healthy historical data with the procedure discussed in Section 2.3.1.

The objective of the condition recognition system is only to infer whether the features are from a healthy machine or not i.e. a novelty is detected and therefore it is unnecessary to use Bayes' rule given by Equation (5). Hence, the GMM of the healthy features and the decision rule given by Equation (7) are used in the condition recognition system to label the data as healthy (i.e. class label of 1) or as a novelty (i.e. class label of 0).

The next step is to evaluate whether the trained condition recognition system can infer the condition of the gearbox correctly for different gear conditions. The testing measurements from the four datasets in Table 3 are used to evaluate the performance of the condition recognition system, with the class labels presented in Figure 15. The healthy gear data in Figure 15(i) and 15(ii) are classified correctly with only three measurements being incorrectly labelled as novelties in Figure 15(i) i.e. they are labelled as class 0. These three measurements are outliers in the dataset.

Most of the initial damaged gear measurements of the first experiment are classified as healthy in Figure 15(iii) due to the small damage which resulted in an overlap in the healthy feature region as seen in Figure 14(ii). As the condition of the gear deteriorates, novelties are detected at an increasing rate until the gear tooth ultimately fails. In contrast to the result in Figure 15(iii), the damaged gear of the second experiment does not contain evidence of a long transition from healthy to damage period in Figure 15(iv). A few of the measurements are classified as novelties which can indicate the presence of damage, with the failure occurring in the final stages of the experiment.

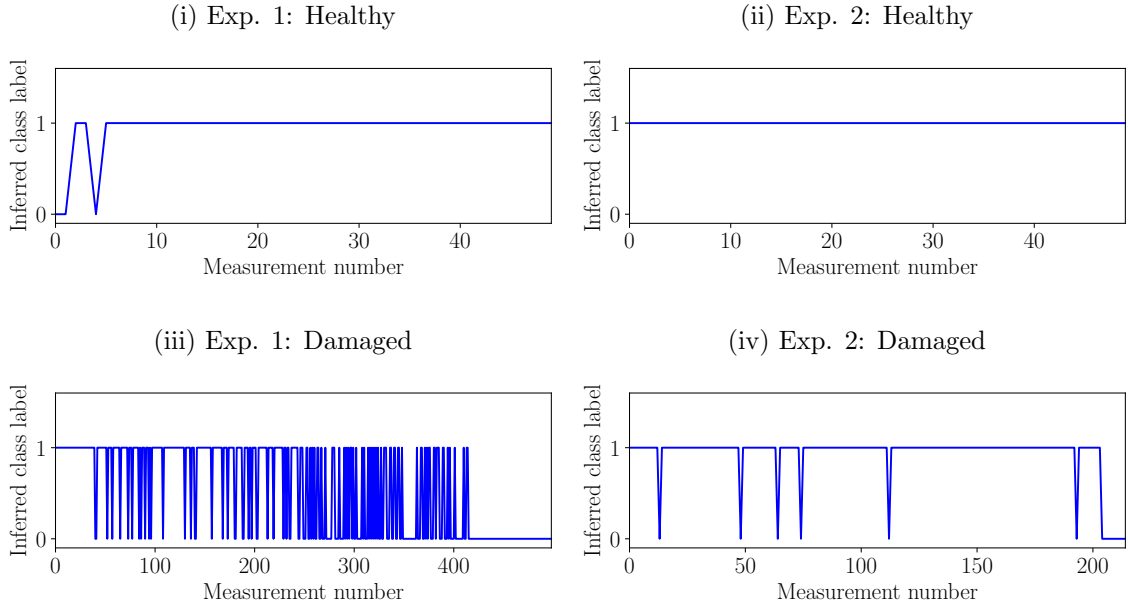


Figure 15: The output of the condition recognition system in Figure 2 for the different datasets when only healthy historical data are available. The inferred class label correspond to the class labels in Figure 2, with 0 indicating a novelty and 1 corresponding to the healthy class.

The detection of the failure in the final stages of the experiment is not a deficiency of the method; the seeded damage was more severe for the second experiment than for the first experiment which resulted in a significantly shorter experimental time (i.e. 3 days as opposed to 20 days) due to the gear tooth failing abruptly.

The synchronous averaged discrepancy signals, corresponding to the measurements where novelties are detected in Figure 15(iii) and Figure 15(iv), are presented in Figure 16. The many

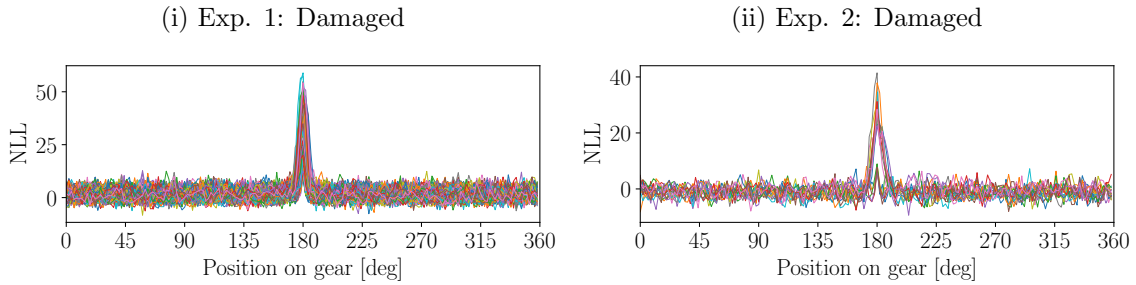


Figure 16: The synchronous averaged discrepancy signals which correspond to the novelties detected in Figure 15(iii) and 15(iv) are presented in Figure 16(i) and Figure 16(ii), respectively. Different colours are used to emphasise that the synchronous averages of different measurements are presented.

synchronous averages in Figure 16(i) make it difficult to interpret the results in the figure and therefore a three-dimensional figure is presented in Figure 17 of the same results. There are not many novelties detected in Figure 15(iv) and not many synchronous averages presented in Figure

16(ii), which makes the results in Figure 16(ii) easy to interpret and therefore a three-dimensional surface is not shown for that data. The results in Figures 16 and 17 indicate that the novelties, detected in Figure 15(iii) and Figure 15(iv), are due to localised damage on a gear tooth. Hence,

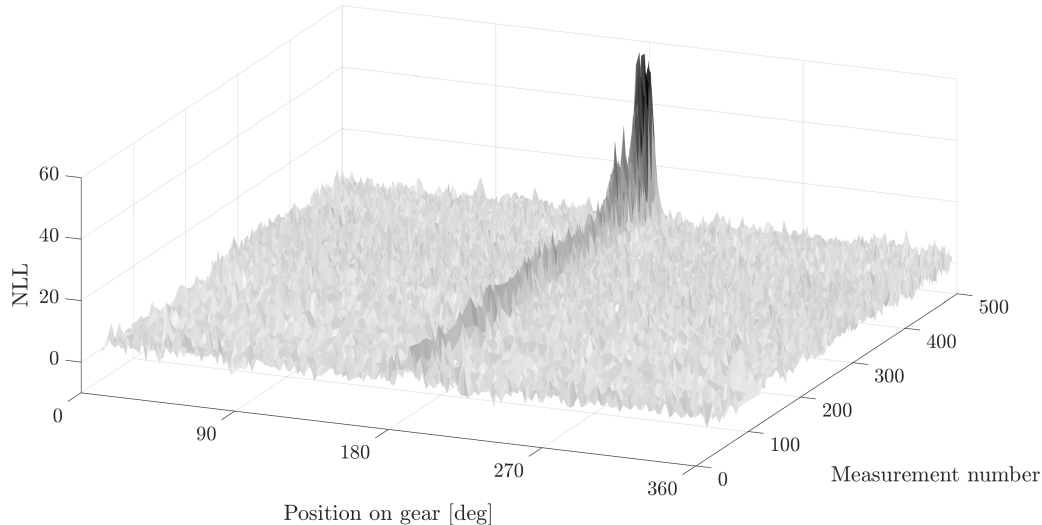


Figure 17: The results in Figure 16(i) are presented as a three-dimensional surface to make it easier to interpret the results.

it is not only possible to automatically label the measurements from a damaged gearbox under the varying operating conditions in Figure 10 as novelties with the condition recognition system, but also to characterise the damage by manually investigating the corresponding synchronous averaged discrepancy signal.

In the next section, it is investigated whether it is possible to update the condition recognition system after historical fault data become available. The ability of the condition recognition system to correctly infer the condition of the gearbox is investigated as well.

4.2.4. Classification with healthy and damaged gear historical data

In this section, it is assumed that the historical damaged gear data, presented in Section 4.1 and Table 3, are available after the condition recognition system has already been optimised as a novelty detector in the previous section. Hence, the historical fault data, labelled as the training data in Table 3 need to be incorporated into the condition recognition system. The training data of the damaged class are spaced throughout the life of the gear from its initial damaged state and its failure state as described in Table 3.

The condition recognition system is updated by firstly optimising a new GMM on the damaged

class discrepancy signal features, whereafter the prior probabilities of the healthy and damaged gear classes are calculated as discussed in Section 2.3.1. The resulting prior probabilities of the classes are therefore $P(C_{healthy}) = 2/3$ and $P(C_{damaged}) = 1/3$. Hence, the condition inference procedure works in two steps: Firstly, the likelihood $p(\mathbf{r}|C_i)$ of the GMM of the damaged gear (optimised in this section) and the GMM of the healthy gear (optimised in the previous section) are calculated for the discrepancy signal features of a new measurement. Thereafter, the likelihoods and the prior probabilities are used with Bayes' rule to infer the condition of the gear as described in Section 2.3.1. The inferred class is then either accepted or rejected (i.e. labelled as a novelty) according to the procedure in Section 2.3.2.

The OSR GMM using Bayes' rule, with and without the decision rule given by Equation (7), are investigated on the four experimental datasets, with the resulting decision boundaries presented in Figure 18. Both classifiers distinguish the different classes well, but the classifier without the

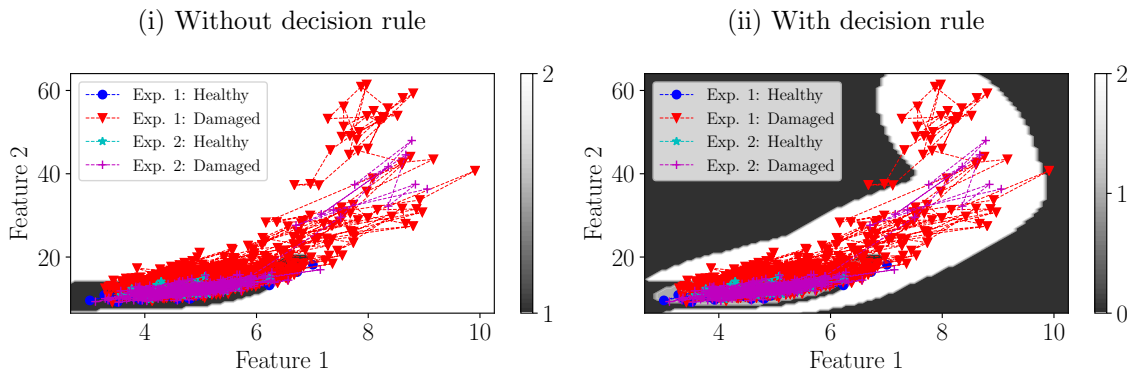


Figure 18: The decision region for the GMM condition recognition system in Section 2.3.1 with and without the decision rule is superimposed on the features of the experimental datasets. Note: Different colour bar ranges are used in Figure 18(i) and 18(ii)

decision rule is incapable of detecting new classes and outliers as evidenced in Figure 18(i). The classifier with the decision rule learned the trajectory of the data as the gear's condition deteriorates and can indicate when there is a region in the feature space unsupported by the training data as seen in Figure 18(ii).

The posterior probability of the different classes calculated using Equation (5) for the four datasets, are smoothed using a moving average window with a length of 5% the number of measurements and a 90% overlap, and presented in Figure 19. The healthy gear data in Figure 19(i) and Figure 19(ii) are classified correctly. The damaged gear posterior probabilities in Figure 19(iii) and Figure 19(iv) indicate that the gears started from approximately healthy conditions, whereafter the

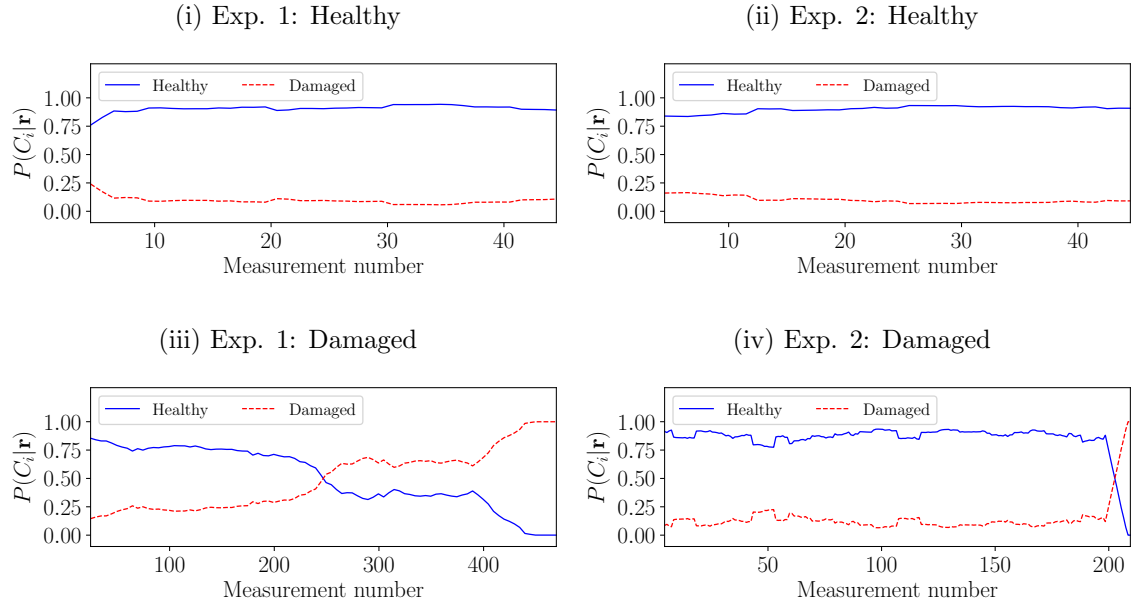


Figure 19: Posterior probability of the two classes for the four experimental datasets using the condition recognition system with the decision rule.

damage progressed up to the failure of the tooth. The damage in Figure 19(iii) can be detected at the 244th measurement, while the damage in Figure 19(iv) is detected at the 203rd measurement. The different characteristics in Figure 19(iii) and Figure 19(iv) are attributed to the different characteristics of the seeded damage and experimental time discussed in Section 4.1.

The change in probability for the damaged gears is evidence that the condition of the gears change with measurement number, however, it is shown in Section 3 that the hidden states contain useful information as well. The hidden states of the GMM for the associated class, determined by the procedure described in Bishop [45], are presented in Figure 20 for the damaged gears. The two

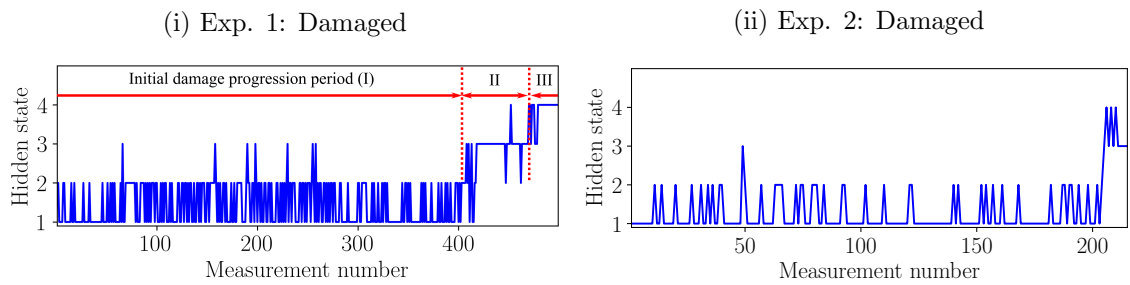


Figure 20: The hidden state in which the data reside as a function of measurement number for the damaged gears. Three stages are identified in Figure 20(i) and labelled with Roman numerals.

damaged datasets exhibit very similar characteristics. The first two states are visited for the initial measurements and as the gear deteriorates, state 3 and 4 become more active. The presence of

Stage II and III in Figure 20(i) indicate that the tooth failed or that failure is imminent. Due to the abrupt failure of the second gear tooth, the progression between Stage II and III is not seen. The hidden states can therefore be used in the maintenance decision making process, for example if the gear is classified as damaged and if state 2 is visited more frequently, a warning is issued and if state 3 and 4 (i.e. Stage II to III) is reached, the machine needs to be stopped as soon as possible.

It is concluded that the methodology is capable of detecting novelties and classifying the data correctly. The benefits of the open set recognition framework is clearly highlighted on the synthetic and experimental data. However, the benefits of using discrepancy analysis as a pre-processor for the classification problem has not yet been investigated. This investigation is performed in the next section.

4.3. Comparison to a Conventional Classification Procedure (CCP)

The aim of the investigation in this section is to highlight the benefits of using discrepancy analysis prior to extracting the features used for classification. The feature extraction procedure in Section 2.1.1 is extended for a Conventional Classification Procedure (CCP) to circumvent the use of discrepancy analysis in the condition inference procedure. The following process is used:

- Calculate the CWT of the order tracked vibration signal for the first three gear mesh components according to Section 2.1.1.
- Calculate the statistics given in Table 1 for the entire measurement. Hence, the windowing procedure is not used and this results in a single 30 dimensional feature set to be extracted for each measurement of the four datasets.
- The condition recognition system cannot be efficiently optimised due to the high dimensionality of the features, hence the dimensionality of the features is reduced with Principal Component Analysis (PCA). PCA is used, because it performed the best of all dimensionality reduction methods used for gear diagnostics in Ref. [18]. The PCA model is optimised on the historical data which is used to project the features on a lower-dimensional principal component feature space. The accumulative contribution rate is used to determine the appropriate dimensionality of the principal component feature space [38], which resulted in three dimensions to be used.

In Figure 21, the proposed discrepancy features and the CCP features are compared for 44 historical healthy measurements and 44 historical damage measurements, with the latter spaced

evenly over two days before the first gear failed. The 44 historical damage measurements used in Figure 21(ii) are not included in Table 3 and are used to compare the sensitivity of the features to faults. The two classes in Figure 21(i) are fairly well separated, with only a few samples overlap-

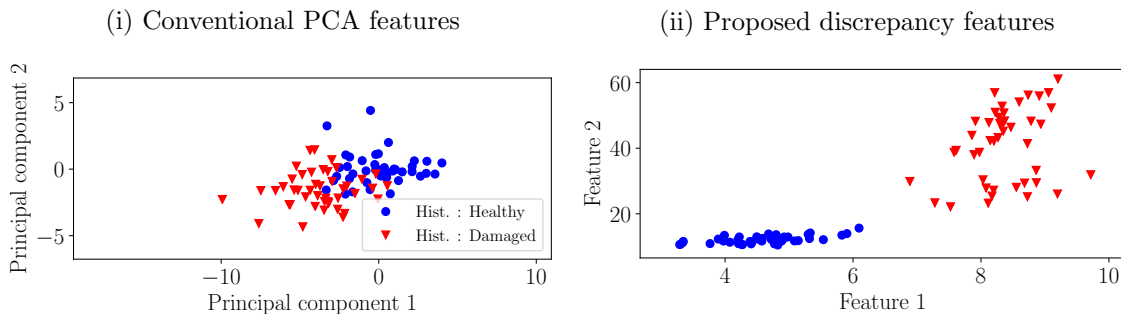


Figure 21: A comparison of the feature space using the proposed discrepancy features in Section 2.1.1 and the conventional features discussed in Section 4.3.

ping. However, the proposed discrepancy analysis features in Figure 21(ii) separate the two classes significantly better than using the principal components of the raw features, which indicate that the proposed discrepancy features are more sensitive to damage. The large variance of the historical damaged data in Figure 21(ii) is attributed to features’ sensitivity to the large change in machine condition as the gear tooth failed.

In Section 4.2, 44 healthy measurements are used for optimising the healthy GMM and 166 damaged gear measurements are used for optimising the model of the damaged GMM as outlined in Table 3. However, the 166 damaged measurements did not work very well with the PCA model due to the large overlap between the healthy historical data and the healthy portion of the historical fault data as shown in Figure 22(i). In contrast, the condition recognition system in Figure 22(ii) performs significantly better when using the 44 damaged measurements in Figure 21(ii). Performing dimensionality reduction with the complete dataset did not separate the classes well which resulted in poor decision boundaries. Hence, the historical features presented in Figure 21(i) are used in the CCP for all subsequent analyses.

The principal component feature space, superimposed on the decision boundaries of the condition recognition system, is presented in Figure 23 for the four datasets. It is difficult to see the trajectory of the features as the gear condition deteriorates for the two experiments in contrast to Figure 14. There a difference in the trajectory in the final stages of the condition transition process for the two experiments i.e. Exp. 1: Damaged and Exp. 2: Damaged in Figure 14. This resulted in some measurements of the second experiment with the damaged gear i.e. Exp. 2: Damaged to be

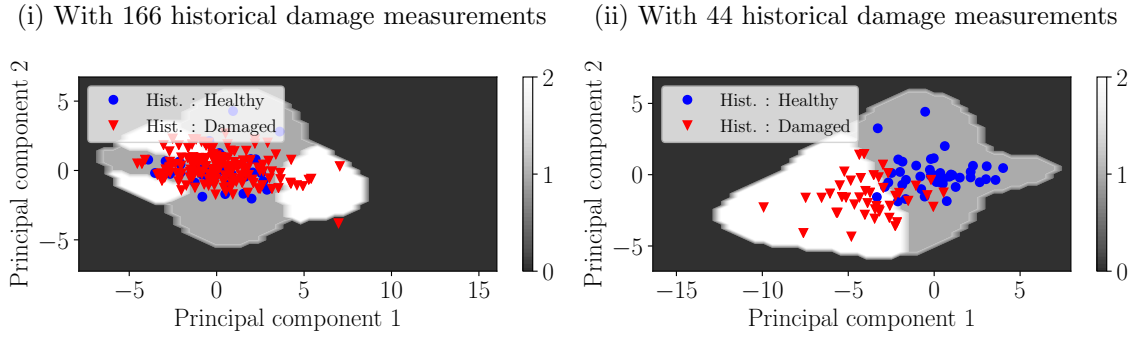


Figure 22: The decision region for the proposed condition recognition procedure with the 44 measurements in Figure 21(i) used for optimising the damaged model in contrast to using the damaged measurements discussed in Table 3. A class label of 0 corresponds to a novelty, a class label of 1 corresponds to a healthy gear and 2 corresponds to a gear with localised damage.

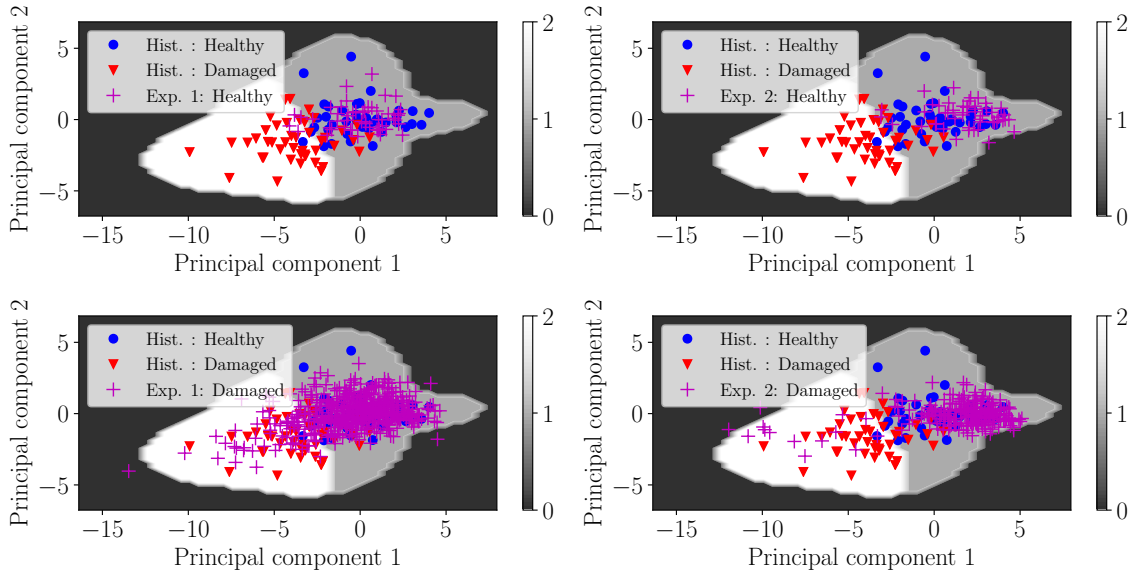


Figure 23: The principal component feature space of the features from the four experimental datasets using the procedure discussed in Section 4.3 is shown with the decision regions of the condition recognition system.

incorrectly classified as outliers.

The posterior probability of the datasets, calculated with the same procedure as section 4.2, is presented in Figure 24. It is clear from the results that the classifier is capable of predicting the class label correctly for the healthy measurements and in the final measurements of the damaged gears, the damage is detected. The damage in Figure 24(iv) is detected at the same measurement number as in Figure 19(iv) due to the fact that the tooth failed abruptly. However, the damage in Figure 24(iii) is detected at the 460th measurement, which is significantly later than in Figure 19(iii), because the discrepancy features are more sensitive to changes in machine condition.

Hence, processing the data with discrepancy analysis results in a better performance (e.g. it

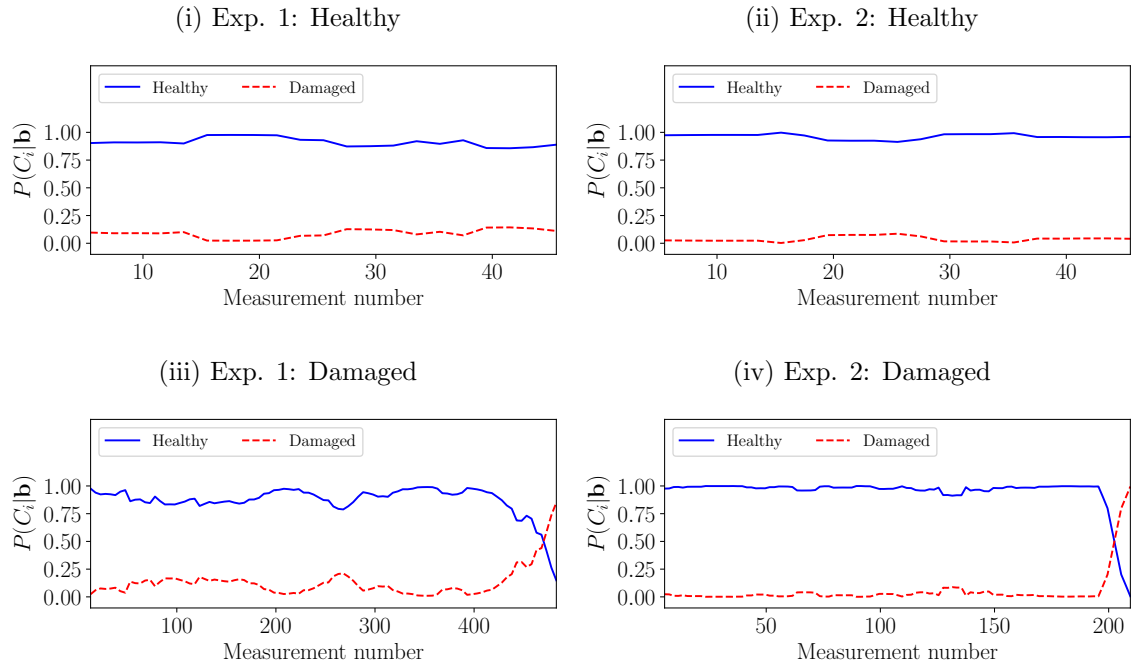


Figure 24: The posterior probabilities of the two classes for the four experiments using the principal component features with the condition recognition system, i.e. discrepancy analysis is not used.

detects damage earlier and clearer decision boundaries are obtained) and therefore it is sensible to include discrepancy analysis in the OSR methodology.

5. Conclusion and recommendations

In this paper, an open set recognition methodology is proposed and investigated on a synthetic dataset and on experimental datasets acquired under varying operating conditions. It is shown that the methodology is capable of detecting novelties when only historical fault data are available, it is able to predict the condition of the machine from the data if historical fault data are available, it is easy to adapt the model when historical fault data of a new damage mode become available and class predictions are only made when the data are in regions supported by the historical data. If the new data are not supported by the historical data, a novelty is detected and the data can be manually analysed to infer the condition of the machine. The methodology is also compared to the case where discrepancy analysis is not used to process the data. It is shown that the proposed discrepancy features are very sensitive to damage and they perform significantly better than using principal component analysis to reduce the dimensionality of the feature space. Lastly, it is also shown that learning the trajectory of the data as the condition of the machine changes can improve the predictive capabilities of the classifier, especially in the open set recognition framework.

In future work, it is suggested that this methodology be extended to the case where multiple faults are present, for more complicated gearboxes such as planetary gearboxes and also for other types of faults such as bearing damage. Distinguishing between localised and distributed gear damage is very important and needs to be addressed in future experimental studies as well. It is suggested to investigate and compare other potential feature extraction methods and models for OSR. The predictive uncertainties obtained from Bayesian data analysis can also potentially be used in the condition monitoring field for OSR problems. Deep learning techniques, which have much potential for rotating machine diagnostics, must also be adapted to incorporate the open set recognition problem and the continuous transitions between healthy and damaged states in the model. Lastly, the experimental setup, the operating conditions and the seeded damage are simplifications of actual characteristics of gearboxes found in industrial environments and therefore the methodology needs to be validated on industrial machines as well.

Acknowledgements

The authors gratefully acknowledge the support that was received from the Eskom Power Plant Engineering Institute (EPPEI) in the execution of the research.

References

- [1] A. K. S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing* 20 (7) (2006) 1483–1510.
- [2] Y. Lin, L. Tu, H. Liu, W. Li, Fault analysis of wind turbines in China, *Renewable and Sustainable Energy Reviews* 55 (2016) 482–490.
- [3] J. Antoni, Cyclostationarity by examples, *Mechanical Systems and Signal Processing* 23 (4) (2009) 987–1036.
- [4] Z. K. Peng, F. L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography, *Mechanical Systems and Signal Processing* 18 (2) (2004) 199–221.
- [5] Y. Lei, J. Lin, Z. He, M. J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, *Mechanical Systems and Signal Processing* 35 (1-2) (2013) 108–126.

- [6] D. Abboud, J. Antoni, S. Sieg-Zieba, M. Eltabach, Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment, *Mechanical Systems and Signal Processing* 84 (2017) 200–226.
- [7] D. Abboud, J. Antoni, Order-frequency analysis of machine signals, *Mechanical Systems and Signal Processing* 87 (2017) 229–258.
- [8] J. Antoni, The infogram: Entropic evidence of the signature of repetitive transients, *Mechanical Systems and Signal Processing* 74 (2016) 73–94.
- [9] J. Urbanek, T. Barszcz, R. Zimroz, J. Antoni, Application of averaged instantaneous power spectrum for diagnostics of machinery operating under non-stationary operational conditions, *Measurement* 45 (7) (2012) 1782–1791.
- [10] I. Javorskyj, I. Kravets, I. Matsko, R. Yuzefovych, Periodically correlated random processes: Application in early diagnostics of mechanical systems, *Mechanical Systems and Signal Processing* 83 (2017) 406–438.
- [11] X. Xu, Z. Qiao, Y. Lei, Repetitive transient extraction for machinery fault diagnosis using multiscale fractional order entropy infogram, *Mechanical Systems and Signal Processing* 103 (2018) 312–326.
- [12] X. Y. Wang, V. Makis, M. Yang, A wavelet approach to fault diagnosis of a gearbox under varying load conditions, *Journal of Sound and Vibration* 329 (9) (2010) 1570–1585.
- [13] L. Cui, X. Gong, J. Zhang, H. Wang, Double-dictionary matching pursuit for fault extent evaluation of rolling bearing based on the Lempel-Ziv complexity, *Journal of Sound and Vibration* 385 (2016) 372–388.
- [14] L. Cui, N. Wu, C. Ma, H. Wang, Quantitative fault analysis of roller bearings based on a novel matching pursuit method with a new step-impulse dictionary, *Mechanical Systems and Signal Processing* 68 (2016) 34–43.
- [15] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, *Mechanical Systems and Signal Processing* 73 (2016) 303–315.

- [16] C. Li, R. V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R. E. Vásquez, Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals, *Mechanical Systems and Signal Processing* 76-77 (2016) 283–293.
- [17] X. Qi, Z. Yuan, X. Han, Diagnosis of misalignment faults by tachless order tracking analysis and RBF networks, *Neurocomputing* 169 (2015) 439–448.
- [18] X. Wan, D. Wang, P. W. Tse, G. Xu, Q. Zhang, A critical study of different dimensionality reduction methods for gear crack degradation assessment under different operating conditions, *Measurement* 78 (2016) 138–150.
- [19] M. Y. Asr, M. M. Etefagh, R. Hassannejad, S. N. Razavi, Diagnosis of combined faults in Rotary Machinery by Non-Naive Bayesian approach, *Mechanical Systems and Signal Processing* 85 (2017) 56–70.
- [20] H. Ocak, K. A. Loparo, HMM-based fault detection and diagnosis scheme for rolling element bearings, *Journal of Vibration and Acoustics* 127 (4) (2005) 299.
- [21] Q. Miao, V. Makis, Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models, *Mechanical Systems and Signal Processing* 21 (2) (2007) 840–855.
- [22] M. Cerrada, R. V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. Valente de Oliveira, R. E. Vásquez, A review on data-driven fault severity assessment in rolling bearings, *Mechanical Systems and Signal Processing* 99 (2018) 169–196.
- [23] L. Jedliński, J. Jonak, Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform, *Applied Soft Computing* 30 (2015) 636–641.
- [24] F. Jia, Y. Lei, L. Guo, J. Lin, S. Xing, A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines, *Neurocomputing* 272 (2018) 619–628.
- [25] F. Jia, Y. Lei, N. Lu, S. Xing, Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization, *Mechanical Systems and Signal Processing* 110 (2018) 349–367.

- [26] T. Heyns, S. J. Godsill, J. P. De Villiers, P. S. Heyns, Statistical gear health analysis which is robust to fluctuating loads and operating speeds, *Mechanical Systems and Signal Processing* 27 (1) (2012) 651–666.
- [27] D. Fernández-Francos, D. Martínez-Rego, O. Fontenla-Romero, A. Alonso-Betanzos, Automatic bearing fault diagnosis based on one-class ν -SVM, *Computers & Industrial Engineering* 64 (1) (2013) 357–365.
- [28] M. Timusk, M. Lipsett, C. K. Mechefske, Fault detection using transient machine signals, *Mechanical Systems and Signal Processing* 22 (2008) 1724–1749.
- [29] A. E. Lazzaretti, D. M. J. Tax, H. Vieira Neto, V. H. Ferreira, Novelty detection and multi-class classification in power distribution voltage waveforms, *Expert Systems with Applications* 45 (2016) 322–330.
- [30] W. J. Scheirer, A. De Rezende Rocha, A. Sapkota, T. E. Boulton, Toward open set recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (7) (2013) 1757–1772.
- [31] R. Casimir, E. Boutleux, G. Clerc, A. Yahoui, The use of features selection and nearest neighbors rule for faults diagnostic in induction motors, *Engineering Applications of Artificial Intelligence* 19 (2) (2006) 169–177.
- [32] W. J. Scheirer, L. P. Jain, T. E. Boulton, Probability models for open set recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 (11) (2014) 2317–2324.
- [33] M. D. Scherrek, B. D. Rigling, Open set recognition for automatic target classification with rejection, *IEEE Transactions on Aerospace and Electronic Systems* 52 (2) (2016) 632–642.
- [34] D. M. J. Tax, R. P. W. Duin, Growing a multi-class classifier with a reject option, *Pattern Recognition Letters* 29 (10) (2008) 1565–1570.
- [35] T. Heyns, P. S. Heyns, R. Zimroz, Combining discrepancy analysis with sensorless signal re-sampling for condition monitoring of rotating machines under fluctuating operations, *Ninth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies* 2 (2) (2012) 52–58.

- [36] T. Heyns, P. S. Heyns, J. P. De Villiers, Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox, *Mechanical Systems and Signal Processing* 32 (2012) 200–215.
- [37] P. S. Heyns, R. Vinson, T. Heyns, Rotating machine diagnosis using smart feature selection under non-stationary operating conditions, *Insight-Non-Destructive Testing and Condition Monitoring* 58 (8) (2016) 1–8.
- [38] S. Schmidt, P. S. Heyns, J. P. de Villiers, A novelty detection diagnostic methodology for gearboxes operating under fluctuating operating conditions using probabilistic techniques, *Mechanical Systems and Signal Processing* 100 (2018) 152–166.
- [39] S. Schmidt, P. S. Heyns, K. C. Gryllias, A discrepancy analysis methodology for rolling element bearing diagnostics under variable speed conditions, *Mechanical Systems and Signal Processing* 116 (2019) 40–61.
- [40] J. Rafiee, M. A. Rafiee, P. W. Tse, Application of mother wavelet functions for automatic gear and bearing fault diagnosis, *Expert Systems with Applications* 37 (6) (2010) 4568–4579.
- [41] D. H. Diamond, P. S. Heyns, A. J. Oberholster, Online shaft encoder geometry compensation for arbitrary shaft speed profiles using Bayesian regression, *Mechanical Systems and Signal Processing* 81 (2016) 402–418.
- [42] Q. Leclère, H. André, J. Antoni, A multi-order probabilistic approach for Instantaneous Angular Speed tracking debris finding of the CMMNO’14 diagnosis contest, *Mechanical Systems and Signal Processing* 81 (2016) 375–386.
- [43] S. Schmidt, P. S. Heyns, J. P. de Villiers, A tacholeless order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process, *Mechanical Systems and Signal Processing* 100 (2018) 630–646.
- [44] Y. Lei, M. J. Zuo, Gear crack level identification based on weighted K nearest neighbor classification algorithm, *Mechanical Systems and Signal Processing* 23 (5) (2009) 1535–1547.
- [45] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

- [46] C. J. Stander, P. S. Heyns, Instantaneous angular speed monitoring of gearboxes under non-cyclic stationary load conditions, *Mechanical Systems and Signal Processing* 19 (4) (2005) 817–835. doi:10.1016/j.ymssp.2004.10.006.
- [47] T. Marwala, U. Mahola, F. V. Nelwamondo, Hidden Markov Models and Gaussian Mixture Models for Bearing Fault Detection Using Fractals, *International Joint Conference on Neural Networks*, Vancouver, Canada (2006) 3237–3242.
- [48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* 12 (2011) 2825–2830.