

Neural network fault diagnosis of a trolling motor based on feature reduction techniques for an unmanned surface vehicle

Proc IMechE Part I:

J Systems and Control Engineering

1–13

© IMechE 2015

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0959651815581095

pii.sagepub.com



Wathiq Abed, Sanjay Sharma and Robert Sutton

Abstract

This article presents a novel approach to the diagnosis of unbalanced faults in a trolling motor under stationary operating conditions. The trolling motor being typically of that used as the propulsion system for an unmanned surface vehicle, the diagnosis approach is based on the use of discrete wavelet transforms as a feature extraction tool and a time-delayed neural network for fault classification. The time-delayed neural network classifies between healthy and faulty conditions of the trolling motor by analysing the stator current and vibration. To overcome feature redundancy, which affects diagnosis accuracy, several feature reduction methods have been tested, and the orthogonal fuzzy neighbourhood discriminant analysis approach is found to be the most effective method. Four faulty conditions were analysed under laboratory conditions, where one of the blades causing damage to the trolling motor is cut into 10%, 25%, half and then into full to simulate the effects of propeller blades being damaged partly or fully. The results obtained from the real-time simulation demonstrate the effectiveness and reliability of the proposed methodology in classifying the different faults faster and accurately.

Keywords

Feature extraction, feature reduction, time-delayed neural network

Date received: 15 May 2014; accepted: 12 March 2015

Introduction

Automatic marine control systems for ships of all sizes have been and are being designed and developed to meet the needs of both the military and civil marine industries. Although modern ship automatic systems are endowed with highly sophisticated subsystems which are expensive, they also possess manual override facilities in case of emergencies and unforeseen occurrences. However, the luxury of such facilities does not exist on board relatively low-cost unmanned surface vehicles (USVs).

USVs are now being employed by the scientific, off-shore and naval sectors to perform a multitude of different tasks with great effect. As a consequence of their success, these sectors are now demanding longer mission lengths coupled with increasingly more vehicle autonomy. With an escalation in autonomy, comes the need for higher reliability in order for them to better cope with unexpected events.

Hence, there is a growing interest in the use of fault detection and diagnostic techniques in USVs. At

Plymouth University, the *Springer* USV has been built and continues to be evolved by the Marine and Industrial Dynamic Analysis Research Group. The USV being designed primarily for undertaking pollutant tracking and environmental and hydrographical surveys in rivers, reservoirs, inland waterways and coastal waters, particularly where shallow waters prevail. The USV was designed as a medium water plane twin hull vessel which is versatile in terms of mission profile and payload. It is approximately 4 m long and 2.3 m wide with a displacement of 0.6 ton.

The propulsion system consists of two propellers powered by a set of 24 V 334 N Minn Kota Riptide transom mounted saltwater trolling motors. In

School of Marine Science and Engineering, University of Plymouth,
Plymouth, UK

Corresponding author:

Wathiq Abed, School of Marine Science and Engineering, University of
Plymouth, Plymouth PL4 8A, UK.

Email: Wathiq.abed@plymouth.ac.uk

common with many USVs, such permanent magnet DC motors provide the means of propulsion owing to their high efficiency, size and weight.¹

Thus, the timely isolation of faults in a motor will ensure the integrity and safety of a vehicle while not adversely affecting the overall system performance. In practice, when undertaking a mission if necessary, a fault detection and diagnosis can be instigated on board a USV while using telemetry to supply its mission control centre with a status report. In this article, a new approach to the diagnosis of unbalanced faults in a trolling motor operating under stationary conditions is presented. With regard to the structure and content of this article, on completion of this introductory material, section 'Approaches to fault diagnosis' provides a summary of current approaches to fault diagnosis, while section 'Experimental set-up' describes the experimental verification of the faults being considered in the study, and in section 'Fault diagnosis process', a novel fault diagnosis process is presented. This is followed by a description of an innovative fault classification procedure based on neural network (NN) architecture in section 'NN architecture for fault classification'. Finally, conclusions are given in section 'Conclusion'.

Approaches to fault diagnosis

Various fault diagnosis techniques such as model-based approach,¹ soft computing techniques² and motor signal analysis³ use stator current and vibration to classify faults. Reduction of human involvement in the diagnosis process has gradually taken place owing the development of modern artificial intelligence (AI) tools. NNs,⁴ fuzzy and adaptive fuzzy logic systems (FLSs)⁵ and expert systems are good candidates for the automation of the diagnostic procedures.⁶

Wavelet packet transforms (WPTs) and NNs have been implemented to diagnose induction motor stator winding short circuit faults. In one such approach, NNs were trained by an improved generic algorithm (GA) using the stator current as a fault indicator.⁷

An NN is an effective motor fault detection method while avoiding the need for a mathematical model. NNs are also less affected by noise and learn the motor early fault detection process, to give accurate solutions to a particular fault. In addition, an NN can recognise patterns even at high noise levels.⁸

With FLS, complexity of the data-driven fuzzy models can be reduced with so-called merging algorithms.^{9,10} Furthermore, fuzzy models offer some sort of interpretability^{11,12} and adapt their parameters and structures over time.¹³

However, NNs are universal approximates and can represent often higher non-linearity degrees contained in the data better than FLS, especially when time-delayed variable features are fed for training. In addition, FLS models tend to a rule explosion, for instance, the number of rules increases exponentially if the

number of variables or fuzzy sets per variable increases, making it complicated to identify the whole model from the knowledge of an expert only.¹⁴

However, a support vector machine (SVM) is an effective tool for motor fault classification based on statistical learning theory, due to its good generalisation abilities.¹⁵ The standard SVM is based on binary classification problems to determine a linear boundary between the two different classes, by maximising the distance of the nearest data to the boundary in each class. However, a SVM requires rigorous tuning of kernel parameters, and the process of optimising generates a large amount of calculation. Furthermore, SVM has high complexity and needs a wide range of data, and the traditional SVM is unusable with dynamic data.¹⁶

Most of the traditional fault diagnosis techniques operate off-line and are not suitable for online prediction and real-time diagnostics. Recently, time series have been widely implemented for prediction and online fault diagnosis. Serdio et al.¹⁷ proposed a method based on both parallel sub-models and time series multi-scale reconstruction for achieving better efficiency. Also, a separate NN can be used to diagnose abnormal conditions.¹⁸

The high-dimensional data tend to redundancy and cannot be separated well to indicate the condition of faults. Widodo and Yang¹⁹ used principal component analysis (PCA) as a feature reduction tool and then applied a SVM as a fault classifier for analysing high-dimensional data. While PCA works well for linear relationship, it has limited ability to deal with non-linear behaviour of the data.²⁰

Linear discriminant analysis (LDA) and orthogonal fuzzy neighbourhood discriminant analysis (OFNDA) can deal with the non-linear patterns. LDA is used for data classification dimensionality reduction and provides better performance than PCA for high-dimensional data.²¹ However, LDA faces the problem of singularities.²² OFNDA techniques are mostly used in medical data analysis²³ and provide better feature reduction performance compared to PCA and LDA.

To the authors' knowledge, there are no accounts in the literature of using the intelligent features of OFNDA for feature reduction in the fault diagnosis of an electrical motor. This article uses OFNDA as new dimensionality reduction approach in a fault diagnosis of a trolling motor and establishes it has better classification accuracy compared to PCA and LDA.

Experimental set-up

Propellers in the trolling motors are durable but not indestructible. Hard surfaces can damage blades partly or fully and can imbalance the operation of a trolling motor causing significant damage to the internal parts. The diagnoses of these faults are thus necessary for the healthy operation of the trolling motors and critical for USV operations. As shown in Figure 1, the faults were

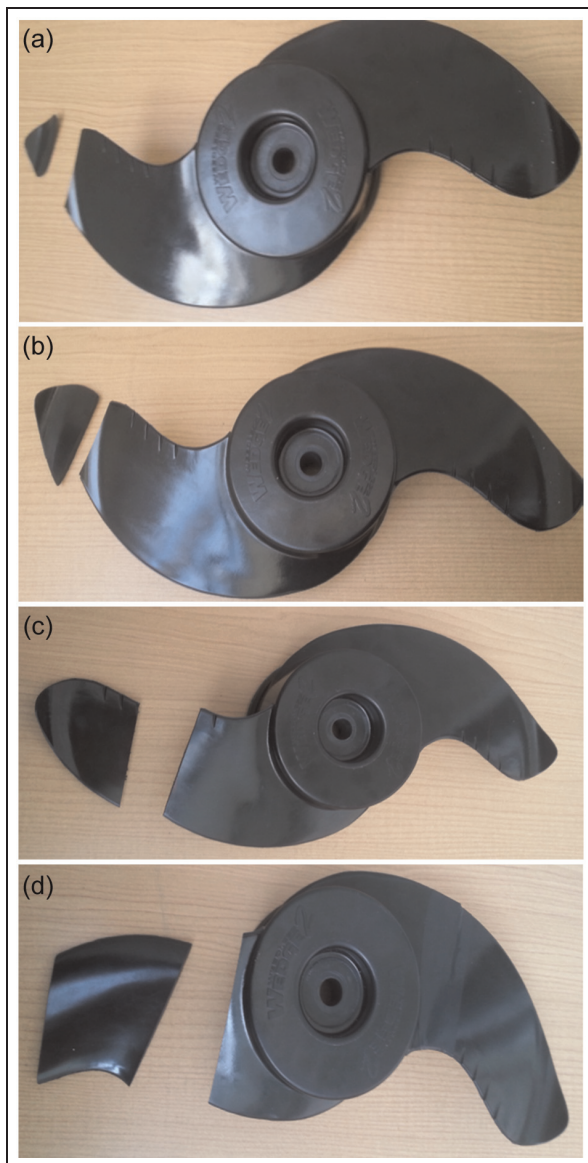


Figure 1. Faults in the blades of the trolling motor: (a) 10%, (b) 25%, (c) half and (d) full cut.

simulated with one of the blades of the trolling motor, which is cut into 10% (F1), 25% (F2), half (F3) and full cut (F4). The proposed technique was used to show the behaviour of the trolling motor under normal operating condition and the four faulty conditions (F1–F4). Three-layered time-delayed neural network (TDNN) is used as fault classifiers, with three output units representing each of the three possible conditions. A heuristic approach was adopted to determine the suitable number of hidden units in the TDNN.

A linear current sensor is used to measure the stator current, and an accelerometer is mounted on the flat surface of the propeller to record the vibration. Sensors' output was logged to a PC via a data acquisition card (NI USB-6009 multifunction I/O device). MATLAB was used to change the duty cycle of the pulse width modulation (PWM) signals, and the motor was powered from 24-V battery supply. Data were gathered for

Table 1. Components used for experiment test.

No.	Item
1	Trolling motor
2	DC power supply
3	Linear current sensor
4	Motor driver
5	Data acquisition system
6	Accelerometer
7	PC

three cases, namely, no fault (normal operating condition), fault F1 and fault F2 at a sampling rate of 3 kHz for a duration of 30 s, and 90,000 sample points were obtained for current and vibration at variable rotor speeds.

Table 1 lists the components used for the trolling motor experimental set-up (Figure 2). A PWM switching technique is used for controlling the speed of the trolling motor by changing the duty cycle of the pulse.

Vibration measurement is the most widely used and effective way to detect faults²⁴ but does not give a clear indication of faults at low speed. Vibration can be picked up from other mechanical parts, thus leading to false positives. Furthermore, because vibration signals are related to all the mechanical elements, they only allow for fault detection rather than fault diagnosis. In order to increase fault diagnosis reliability, especially for critical applications, in addition to vibration, the stator current signal can be used as another fault indicator.

Figure 3 shows the vibration time waveform under four severities of blade fault. It clearly shows large changes in the amplitude at faults, the changes in amplitude in vibration for fault F2 is nearly three times more than fault F1 and indicates a more severe fault in the case where the blade of the motor is F3 and F4. Figure 3 also shows the corresponding current data where variations in the amplitudes are less noticeable compared to the vibration.

The next step in the fault diagnosis will be to use these data to extract features using a wavelet and then reduce the unimportant features using feature reduction techniques. The best selected features are then used to train a NN for fault classification. The total data after feature reduction were divided into three sets: 60% were used for training of the NN, 20% for validation and 20% for testing purposes.

Fault diagnosis process

Figure 4 represents the flow diagram of the fault diagnosis process where the machine represents a trolling motor, and signals such as current and vibration are measured in the signal measurement block. The feature extraction block discrete wavelet transform (DWT) to extract the features from the measured signals and different dimension reduction tools such as PCA, LDA

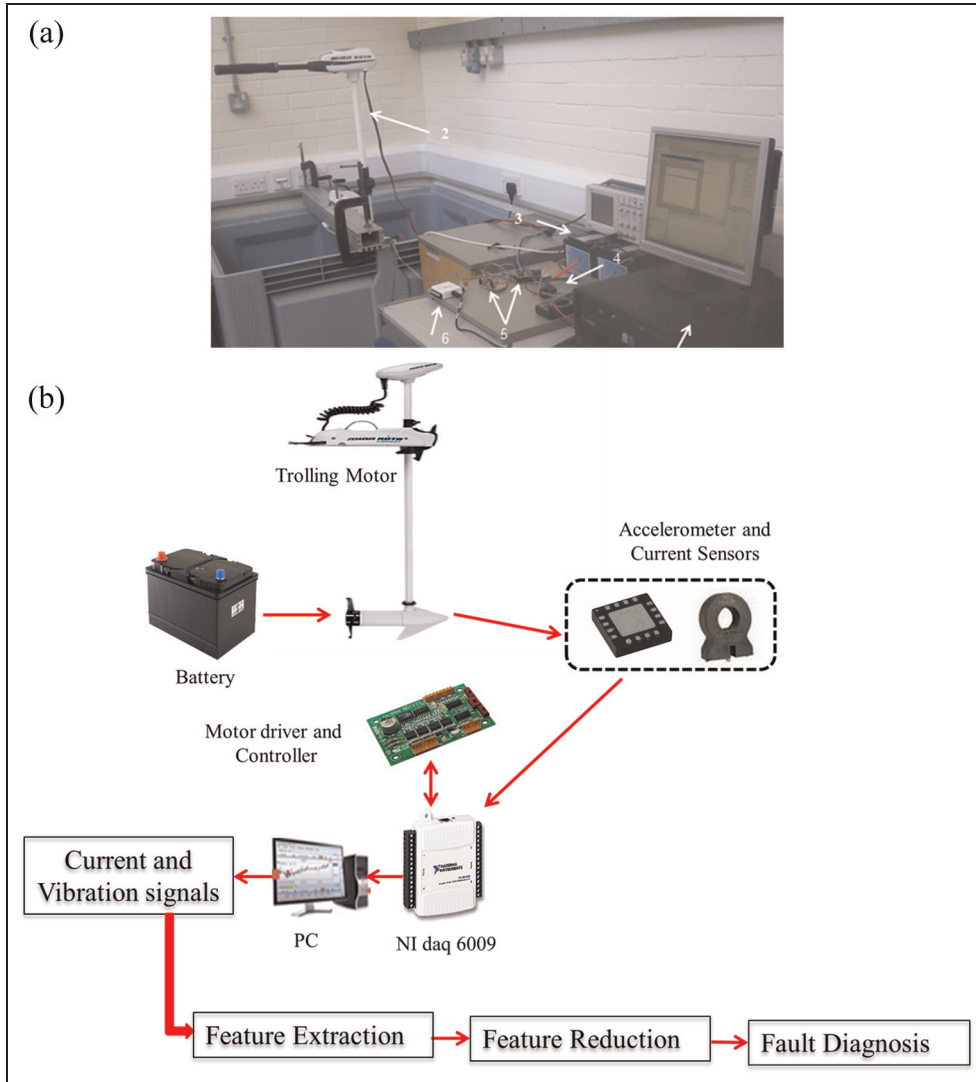


Figure 2. Trolling motor experiment set-up: (a) picture and (b) schematic diagram.

and OFNDA are implemented for selecting the best features. The selected information is then forwarded to a fault decision block where the TDNN classifies the faults.

Feature extraction is usually the first step in any pattern recognition system following the pre-processing step. Extracting the most significant features is crucially important for most pattern recognition problems, motor signal may be processed using three techniques, namely, time-domain analysis, frequency-domain analysis and time–frequency analysis.²⁵

Frequency analysis using fast Fourier transform is not suitable for non-stationary signal. However, short-time Fourier transform (STFT) has limitation related with constant window for all frequencies and is expensive in point view of computational. While wavelet transform (WT) has the ability of exploring signal features with partial characteristic and analysing signals with different time and frequency resolutions.²⁶

Mathematically, the wavelet and scaling functions can be represented as²⁷

$$x(t) = \sum_{n=-\infty}^{\infty} c(n)\varnothing(t-n) + \sum_{n=0}^{\infty} \sum_{n=\infty}^{\infty} d(j,n)2^{j/2}\gamma(2^j t - n) \quad (1)$$

The scaling $c(n)$ and wavelet $d(j,n)$ coefficients for level j are computed as

$$c(n) = \int_{-\infty}^{\infty} x(t)\varnothing(t-n)dt \quad (2)$$

$$d(j,n) = 2^{j/2} \int_{-\infty}^{\infty} x(t)\varnothing(t-n)dt \quad (3)$$

DWT approach is successfully applied to detect and locate faults together with identification of the severity of the faults, and the same approach can be extended to identify the other faults with a significant reduction in the computation time. To emphasise computational efficiency, DWT can achieve such requirement. The DWT can be expressed as²⁸

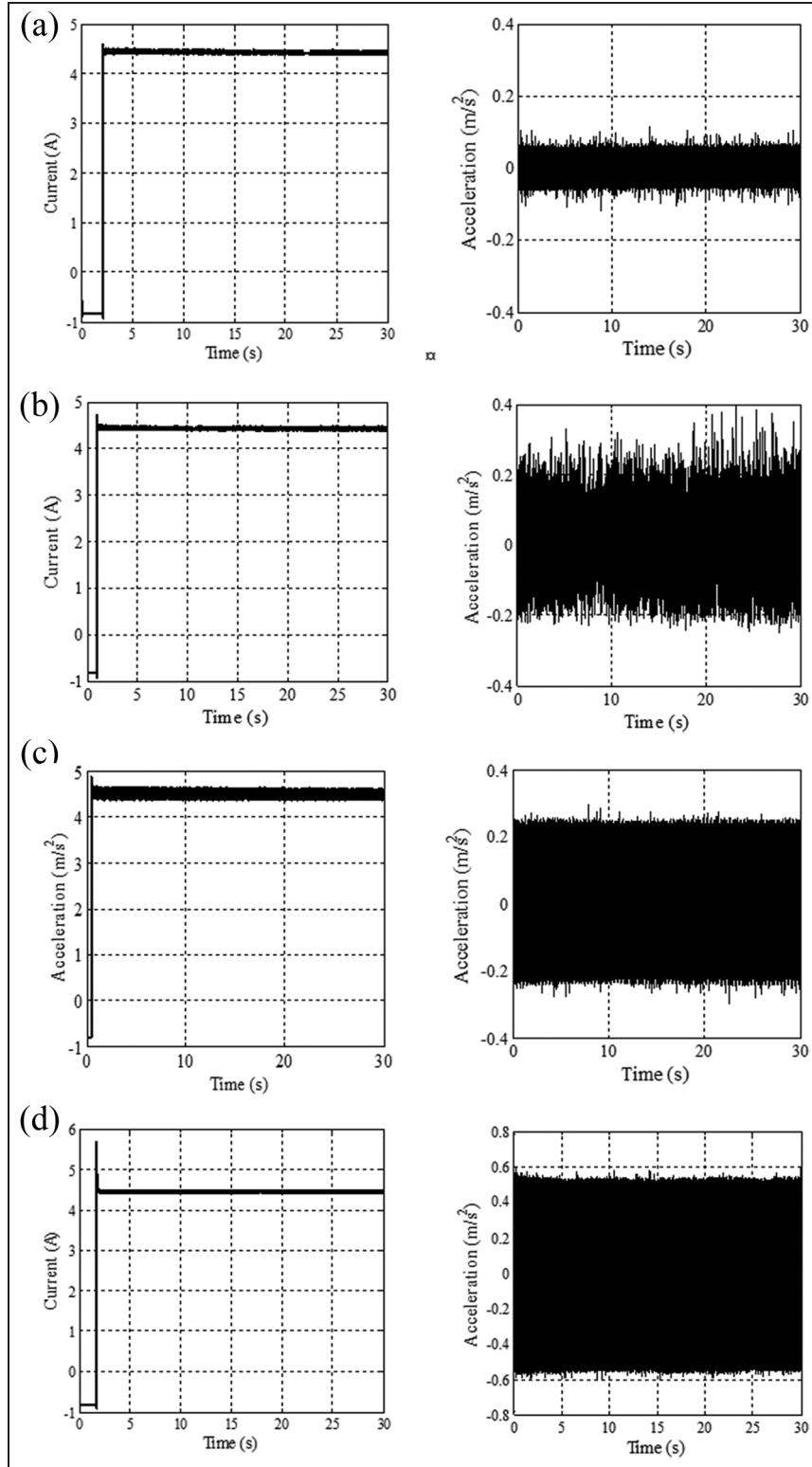


Figure 3. Time domain of the raw vibration (left) and current (right) signals under different operating conditions: (a) F1, (b) F2, (c) F3 and (d) F4.

$$dwt(j, k) = \frac{1}{\sqrt{2^j}} \int x(t) \gamma^* \left(\frac{t - k2^j}{2^j} \right) dt \quad (4)$$

where $x(t)$ is the signal and γ^* is the complex conjugate of the scaled and shifted wavelet function γ . The three-level discrete wavelet decomposition is shown in

Figure 5. At each level, the original signal ($A_o(k)$) is decomposed to separate using low $g[n]$ and high-pass filters $h[n]$ into a detail (d_j) component, which is the high-frequency component, and approximation (a_n), which is the low-frequency component, by correlating the scaled and shifted versions of the wavelet (as in equation (5)).

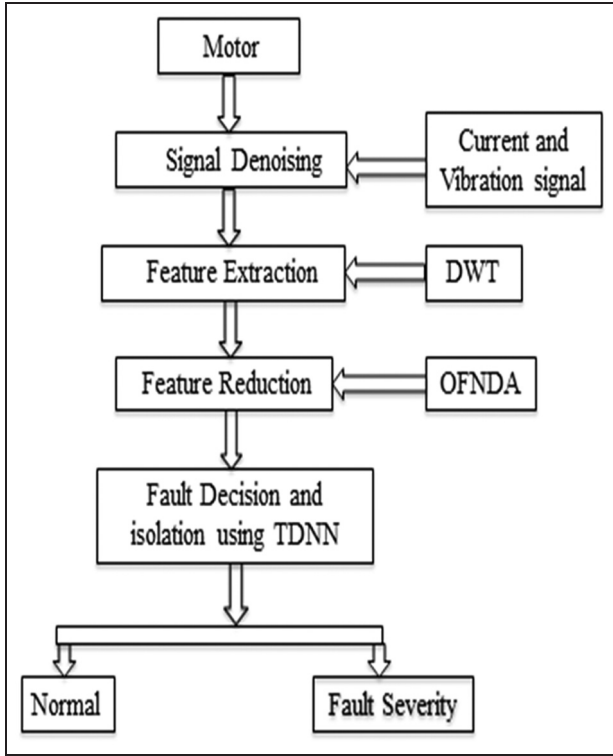


Figure 4. Fault diagnosis process.

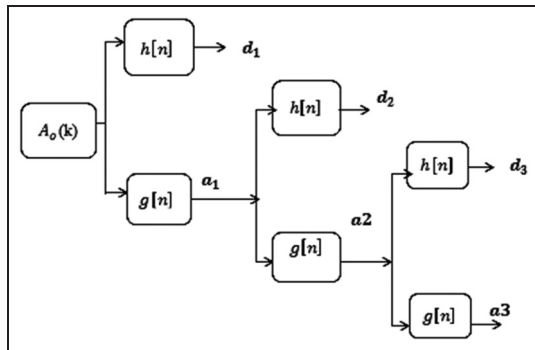


Figure 5. Schematic diagram of discrete wavelet transforms.

The correlation between the signal and the wavelet at each level of scaling and shifting is termed ‘the wavelet coefficient’. The resolution of the signal, which is a measure of the amount of detail of information in the signal, is changed by the filtering operations, and the scale is changed by changing the window size of signals. Resulting from the DWT decomposition, a set of wavelet energy signals (d_j) and (a_j) are obtained²⁹

$$A_o(k) = \sum \vartheta^i \cdot \theta_i^m(t) + \sum_{j=1}^m \sum_i \alpha_i^j \cdot \varphi_i^j = a_n(t) + d_n(t) + \dots + d_1(t) \quad (5)$$

The choice of mother wavelet in DWT is important for better resolution of the signal in time and frequency domains and is selected by trial and error, such as

Deubechies (db), Coiflets (coif) and Symlets (sym). Owing to the use of WT level 5 in this work, six features were obtained for each signal: five details (d1, d2, d3, d4 and d5) and one approximation (a1) using Deubechies (db4) as mother wavelet function.

WT is constituted by different levels. The suitable number of levels of decomposition (n) depends on the sampling frequency of the signal being analysed (f_s). For each one of the proposed approaches, it has to be chosen in order to allow the high-level signals (approximation and details) to cover all the range of frequencies alongside which the sideband component varies during the start.

The minimum number of decomposition levels necessary for obtaining an approximation signal (a) must be established, so that the upper limit of its associated frequency band is under the fundamental frequency (f).³⁰ The data-independent selection (DIS) approach is considered for determining the optimal wavelet level in this work. The DIS approach is based on the following steps³¹

$$2^{-(nL_s + 1)} f_s < f \quad (6)$$

From this condition, the decomposition level of the approximation signal is the integer nL_s

$$nL_s = \text{integer} \left(\frac{\log\left(\frac{f_s}{f}\right)}{\log(2)} \right) \quad (7)$$

where f_s is the sampling frequency (3 kHz), f is the fundamental frequency, and after calculation according to equation (7), scale 5 is selected as the decomposition level in this article.

The two signals (current and vibration) thus provided a total of 12 features. Each combined time–signal (450,000 samples for both current and vibration for normal and four severities of fault) was divided into 3000 windows of 150 data points, so that (3000 × 12) features are obtained and then OFNDA was applied to overcome the irrelevant features and reduced wavelet features to (3000 × 4); finally, OFNDA features are used for training and testing TDNN.

It can be observed from Table 2 that the wavelet energy of the detail coefficients is increased after fault. The most prominent increment of 51% was observed in the d5 level, and this is selected in this article for feature extraction.

The DWT of the signal is shown in Figure 6, and the faults can be ascertained by comparing the wavelet coefficients of the faulty signals with a healthy one.

DWT has been successful in analysing non-stationary signal. However, DWT yields a high-dimensional feature vector as shown in Figure 7. In some cases, the number of features are relatively larger than the number of training samples usually referred as curse of dimensionality and training and testing speed.³²

Table 2. Wavelet energy (details and approximates) under variable load condition (j).

	a5	d1	d2	d3	d4	d5
Healthy	99.97	0.0013	0.0015	0.0038	0.0089	0.0067
F1	99.96	0.0014	0.0016	0.0091	0.0154	0.0118
F2	99.91	0.001	0.0054	0.0281	0.0459	0.0660
F3	99.75	0.0015	0.0056	0.0311	0.0257	0.1780
F4	99.58	0.0016	0.0072	0.0528	0.0845	0.2665

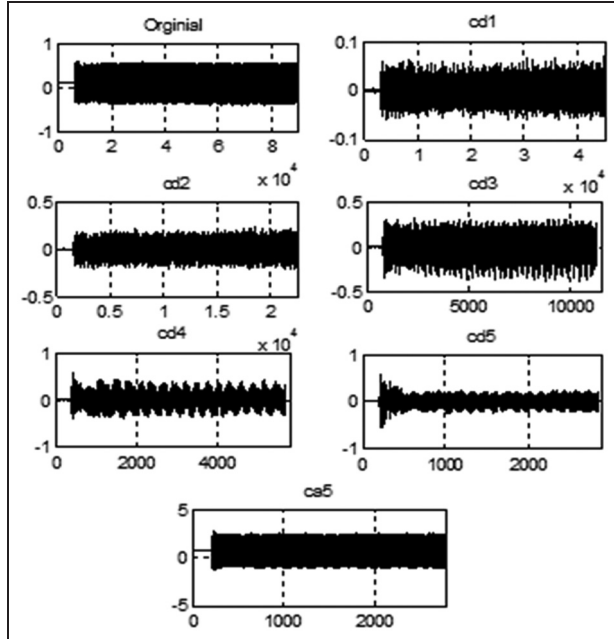


Figure 6. DWT coefficients of faulty case.

In designing a reliable and accurate diagnosis technique, it is critical to find a set of redundant features to reduce additional computational time for classification. An accurate dimensionality reduction tool is thus needed to remove redundant features information.³³

A proper feature reduction technique can reduce the total fault diagnosis time without compromising the quality of classification. A number of techniques such as PCA, LDA and OFNDA are used in feature reduction where the number of features can be reduced using feature selection (FS) or feature projection (FP) techniques.³⁴

PCA is one of the linear feature reduction techniques used to transfer data to a new orthogonal basis whose axes are oriented in the directions of the maximum variance of an input data set. One of the main drawbacks of PCA is that it works to reduce feature redundancy only, without taking into account the relation of features or variables with the specific class labels, and this will affect the classification accuracy.³⁵ PCA also has the drawback of its limited ability to deal with non-linear behaviour of the data.

OFNDA has been recently proposed by Khushaba et al.²³ as a new approach for feature reduction, and it

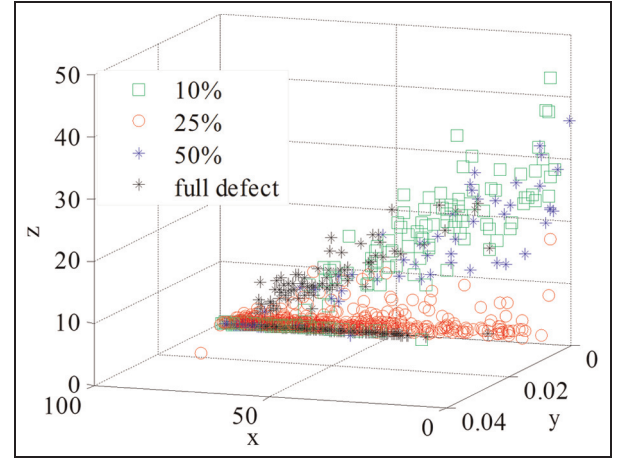


Figure 7. DWT wavelet features for different blade fault severities.

works to maximise the distance between features that belong to different classes (S_b) and minimise the distance between features in the same class (S_w) while taking into account the contribution of the samples to the different classes. OFNDA has been successfully applied to classify four classes of rolling element bearing defects and normal conditions working under variable speed and load conditions.

However, OFNDA has overcome the singularity problems for LDA. In OFNDA process, the first step is to apply PCA to remove any redundancy that may cause singularity before starting discriminant analysis and keep all principal components to prevent loss of any useful information.

Figure 8 illustrates the OFNDA process. Then, the computation of the proposed fuzzy neighbourhood discriminant analysis (FNDA) proceeds by calculating S_w and S_b as given by

$$S_w = \sum_{i=1}^c \sum_{k=1}^{l_i} \mu_{ik} (X_k - X_j)(X_k - X_j)^T \quad (8)$$

$$S_b = \sum_{i=1}^c \mu_{ik} (U_i - X_x)(U_i - X_x)^T \quad (9)$$

$$G_{FNDA} = \arg \text{mat} \text{trac} \left(\frac{G^T S_b G}{G^T S_w G} \right) \quad (10)$$

$$G = G_{FNDA} \cdot G_{PCA} \quad (11)$$

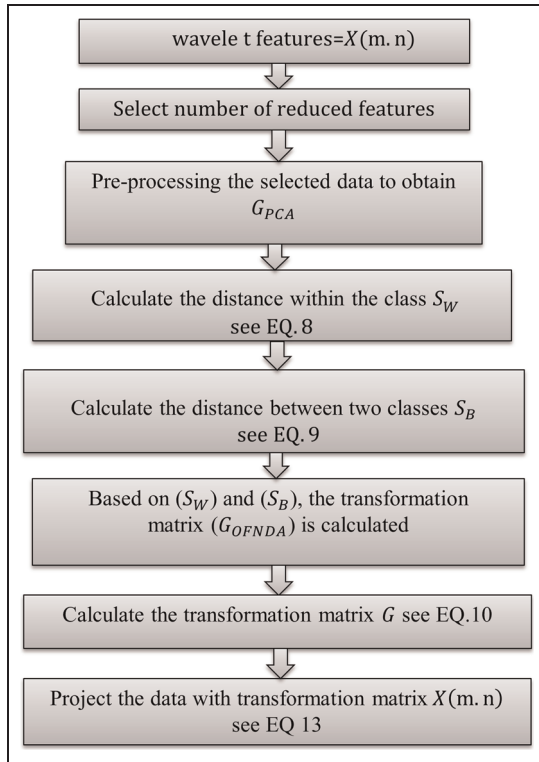


Figure 8. Steps of OFNDA performance.

where μ_{ik} is the membership of pattern k in class i , X_k is the K th sample and U_i is the mean of the patterns that belong to class i .

Given the universal set $X = \{x_1, x_2, \dots, x_l\}$, where x is the feature vector, and $k = 1, 2, \dots, l$ is the number of samples. For simplicity, it will be useful to describe the membership degree that the k th data point has in the i th class by the following equation

$$\mu_{ik} = \mu_i(x) \in [0, 1] \quad (12)$$

X_x is the mean of the training samples, and this is in turn given as follows

$$X_x = \frac{\sum_{i=1}^c \sum_{k=1}^{l_i} \mu_{ik} X_k}{\sum_{i=1}^c \sum_{k=1}^{l_i} \mu_{ik}} \quad (13)$$

$G_{FND A}$ is the transformation matrix related to PCA and OFNDA, respectively

$$X(m \cdot n) = X(m \cdot n) \cdot G_{FND A} \quad (14)$$

LDA is another supervised feature reduction technique which searches for the projection axes on which the distance between data points of different classes are increased, and those between of the same class are minimised. Further details of the LDA can be found in Eleyan and Demirel.³⁶

The main difference between LDA and PCA is that LDA does data classification, whereas PCA does feature classification. In PCA, the shape and location of

the original data sets change when transformed to a different space, while LDA does not change the location but only tries to provide more class separation and draws a decision region between the given classes.

However, OFNDA is implemented to minimise the distance between features in the same class and maximise the distance between features for different classes using the idea of neighbourhood during fuzzy discriminant analysis (FDA). OFNDA differs from conventional Fisher's linear discriminant, that is, a classification method that projects high-dimensional data onto a line and performs classification in this one-dimensional space. The projection maximises the distance between the means of the two classes while minimising the variance within each class and OFNDA mapping data point into a subspace. The above-mentioned feature reduction techniques were able to reduce drastically the number of wavelet features originally from 12 to 4 enabling faster computation.

Figure 9 illustrates the projection of the two reduced features using PCA, LDA and OFNDA for different operating conditions. It can be observed from the graph that boundaries between different operating conditions are more distinct in the case of OFNDA technique.

Table 3 also compares the effectiveness of these three feature reduction techniques for fault classification, and it can be observed from the results that the OFNDA produced better classification accuracy both on validation and on test data compared to PCA and LDA.

NN architecture for fault classification

A three-layered NN, as shown in Figure 10, was used as a fault classifier. The features from OFNDA were fed as input to the NN first without any input time-delay units, then with 5 input time-delay units, and finally with 10 input time-delay units.

In order to classify the three possible conditions, the output layer consisted of three units. Each output unit outputs either a 0 or 1, and thus, the three operating conditions can be encoded as follows: (1 0 0) corresponds to no fault with low severity (F1) and (0 0 1) to an unbalanced load with high severity (F2) fault, where the three binary digits represent the output of the first, second and third units of the output layer, respectively.

In a multilayer perceptron (MLP), NN as in this case, the input to each unit of the hidden (H) and output layers is computed as a weighted sum of the activation of the units in the previous layer. MLP is a feedforward NN model that maps sets of input data onto a set of appropriate outputs. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. The most widely used activation function, and the one used here, is the sigmoid function (non-linear activation function) and is given by

$$f(t) = \frac{1}{1 + e^{-t}} \quad (15)$$

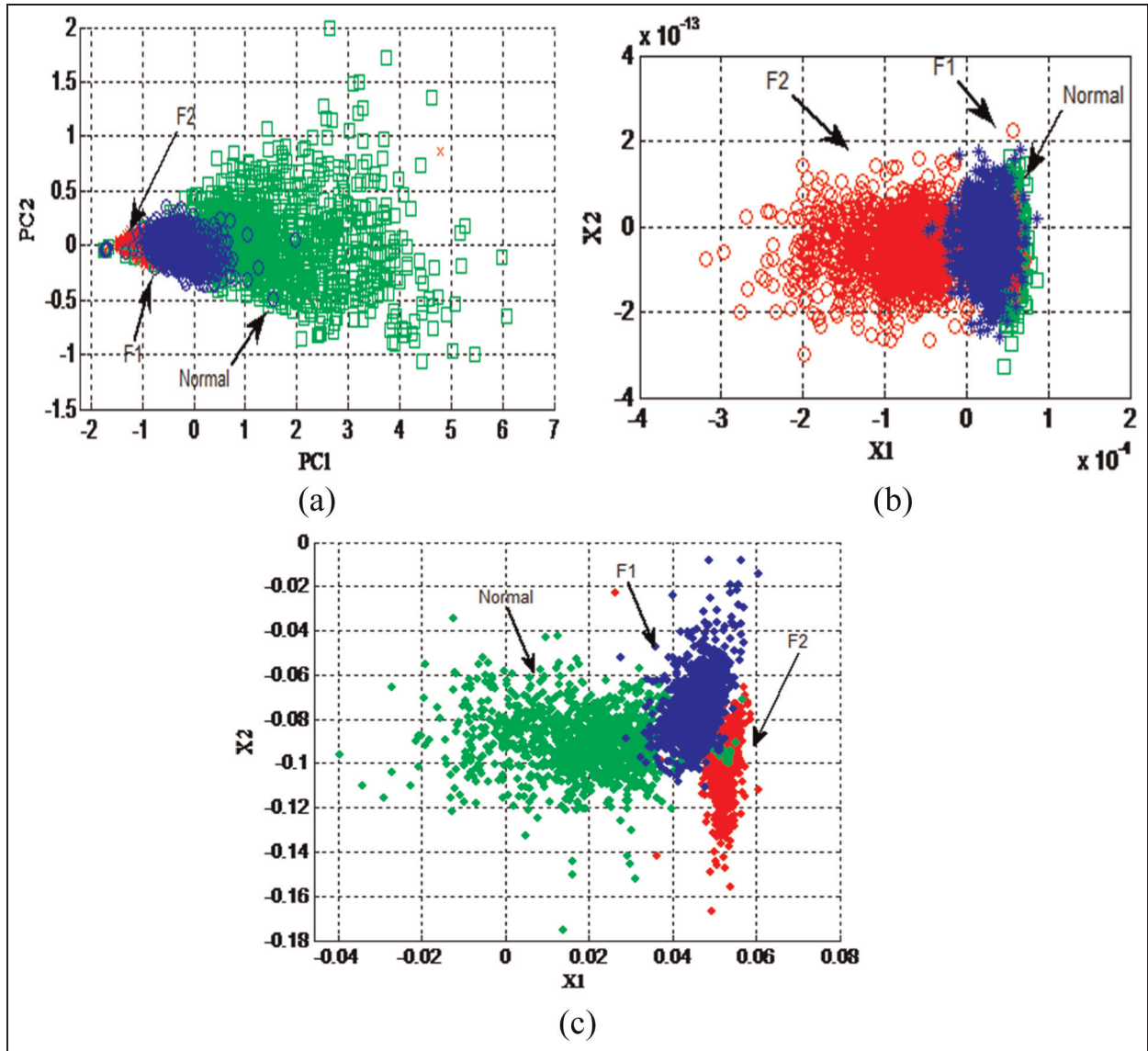


Figure 9. Feature reduction techniques: (a) PCA, (b) LDA and (c) OFNDA.

Table 3. Accuracy of the trained NNs with different combinations of the parameters.

Time delay	Λ	H	Final cost	Validation set (%)	Test set (%)
0	0.01	5	0.962	91.7	98.44
0	0.1	5	0.265	97.6	94.30
0	1	5	0.448	97.2	95.41
0	0.01	10	0.963	91.7	89.72
0	0.1	10	0.265	97.63	94.30
0	1	10	0.453	97.22	95.28
5	0.01	5	0.807	98.60	96.08
5	0.1	5	0.148	99.16	98.88
5	1	5	0.265	99.16	99.02
5	0.01	10	0.181	99.16	99.02
5	0.1	10	0.159	99.16	97.90
5	1	10	0.267	99.16	99.02
10	0.01	5	0.230	93.94	87.04
10	0.1	5	0.131	97.18	96.05
10	1	5	0.241	99.01	98.59
10	0.01	10	0.279	98.87	97.88
10	0.1	10	0.133	97.88	96.16
10	1	10	0.241	99.01	98.59

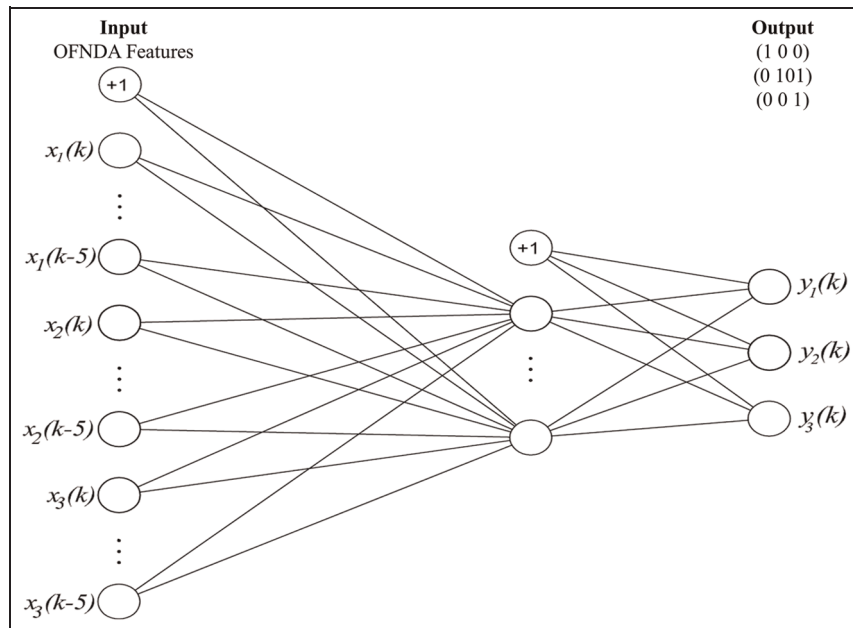


Figure 10. NN architecture.

The training of the NN is the process of adjusting the connection weights between the units in each layer, including those of the bias units. Supervised learning was used to train the NN with a training data set. The process consists of initially assigning some random weights to each of the connections. Forward propagation is then used to obtain the predicted output for each training set sample.

A mean squared error cost function is computed based on the predicted and desired outputs, and the training of the network consists of finding a set of weights that minimise this cost function. Although many techniques are available for minimisation of non-linear functions, such as the gradient descent method, the strategy used here is a slightly more advanced optimisation routine implemented in the function *fmincg* freely available for an Octave.

The optimisation routine was selected to run for 7500 iterations, and the classification accuracy of the trained NN was evaluated. The cost function defined in equation (10) for minimisation incorporated a regularisation term to avoid over-fitting, controlled by a regularisation parameter λ

$$J = \frac{1}{2m} \sum_{k=1}^m (y(k) - \hat{y}(k))^2 + \frac{\lambda}{2m} \sum_{i=1}^n \theta_i^2 \quad (16)$$

where m is the number of sample, $y(k)$ is the target, $\hat{y}(k)$ is the predict value, λ is the regularisation parameter, n is the number of NN weights and θ_i is the NN parameter. Training of the NN involves trying several different values of λ that are used to control the size of NN coefficients based on ridge logistic regression principles, number of input delay units and number of hidden units to achieve the optimised set for different parameters.

For a fast heuristic-based approach, this parameter is set according to minimal and maximal eigenvalues of the regression matrix. The trained NN was evaluated on validation and test data sets (see Figure 13).

Table 3 presents the results with different combinations of the parameters, indicating that better classification accuracy was obtained both on validation and test data sets with five input delay units and five hidden units. Using too few neurons in the hidden layers will result in under-fitting, which help to adequately detect the signals in a complicated data set, whereas using too many neurons in the hidden layers can result in over-fitting.

Over-fitting occurs when the NN has so much information processing capacity that the limited amount of information contained in the training set is not enough to train all of the neurons in the hidden layers. Furthermore, an inordinately large number of neurons in the hidden layers can increase the time it takes to train the network. Several trial-and-error methods are used to optimise the suitable numbers of hidden neuron.

In this article, the above TDNN architecture with five hidden units and with five input delay units was further trained and tested without any feature reduction techniques and with feature reduction techniques using PCA and LDA. The validation and test results, as shown in Table 4, for all cases indicate that feature reduction with OFNDA technique provided better fault classification accuracy compared to without feature reduction, PCA and LDA techniques.

A further test with data obtained under normal and faulty conditions was constructed to simulate the real-time operating behaviour of the trolling motor as shown in Figure 11. The trolling motor was run under normal operating condition for 2.7s and then faults were introduced.

Figure 12 shows the overall fault diagnosis test for a motor operating under different severities of blade fault based on OFNDA features at high speed (Figure 12(a)) and low speed (Figure 12(b)). Figure 13 also shows the correct classification by the proposed technique within

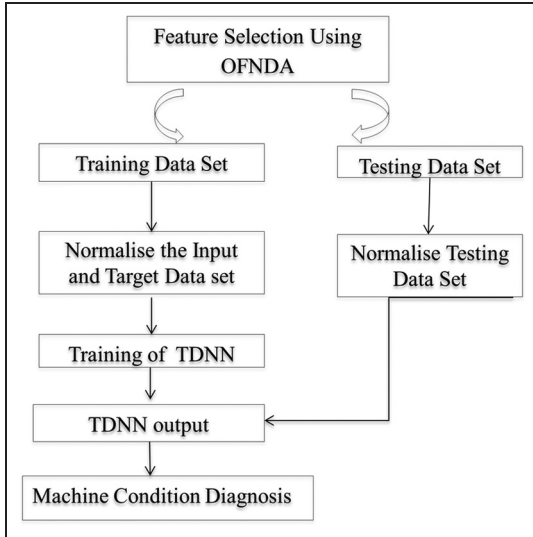


Figure 11. TDNN process.

less than 0.5s of the occurrence of the fault. All the duration times of the misclassifications are less than 0.7s.

In practice, such misclassification times would not be noticeable. Hence, all the misclassifications can be considered spontaneous and can be ignored. Generally, OFNDA shows better fault prediction performance compared with PCA and LDA as shown in Figure 13.

Conclusion

This article proposes a new methodology for fault diagnosis of a trolling motor under two unbalanced load operating conditions. A DWT is used as a feature extraction tool to obtain a better resolution of the signal in the time and frequency domains and then the feature reduction methods PCA, LDA and OFNDA were applied to obtain the best features for fault classification. These features were fed to a TDNN for classifying the faults, and the results showed that better classification accuracy was obtained with OFNDA techniques. Further test simulating the real operating behaviour of the trolling motor under normal and

Table 4. Comparison of the performance of different reduction methods.

TDNN structure	Validation set accuracy%	Test set accuracy (%)
Without dimensional reduction	94.4	97.81
With PCA	98.60	94.60
With LDA	94.72	92.50
With OFNDA	98.15	98.99

TDNN: time-delayed neural network; PCA: principal component analysis; LDA: linear discriminant analysis; OFNDA: orthogonal fuzzy neighbourhood discriminant analysis.

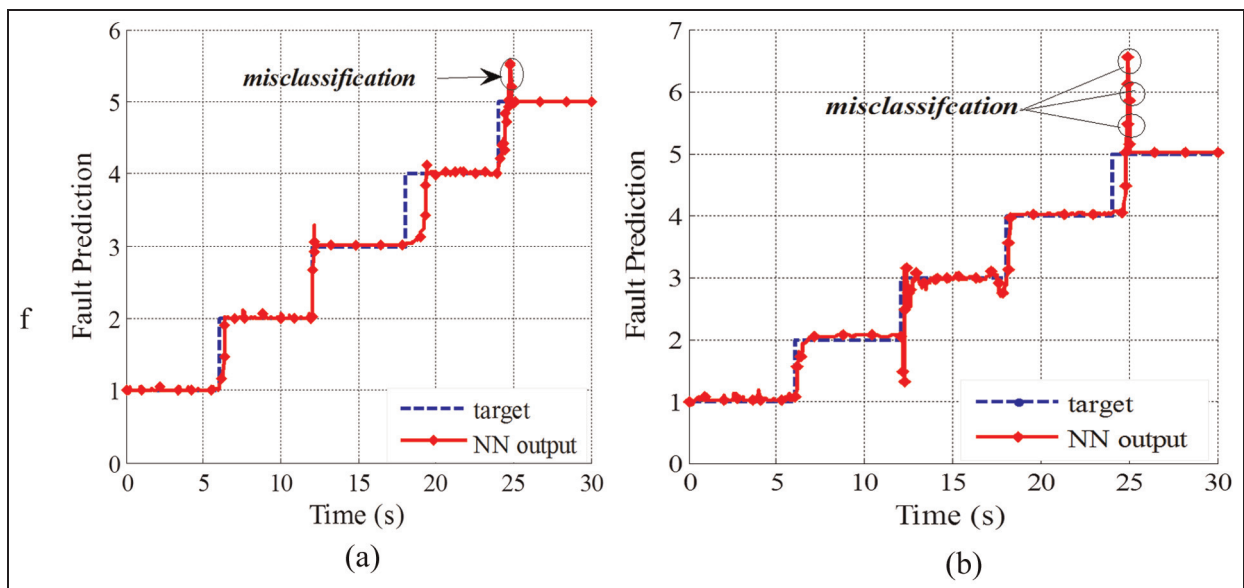


Figure 12. Overall fault diagnosis test for motor operating under different severities of blade fault (F1, F2, F3 and F4) at (a) high speed and (b) low speed.

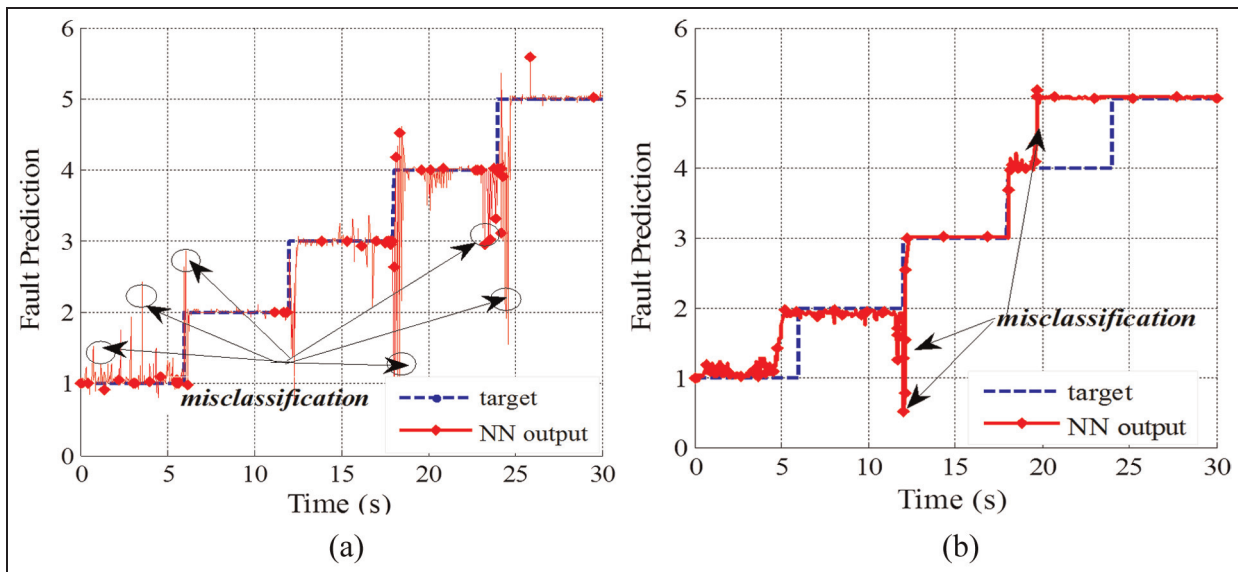


Figure 13. Overall fault diagnosis test for motor operating under different severities of blade fault (F1, F2, F3 and F4) using (a) PCA features and (b) LDA features.

faulty conditions also confirmed the superiority of the proposed method which can easily be applied to real-time classifications on board the *Springer USV*. The proposed technology is portable to other types of autonomous vehicle.

Declaration of conflicting interests

The authors declare that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- Berriri H, Naouar W and Slama I. Easy and fast sensor fault detection and isolation algorithm for electrical drives. *IEEE T Power Electr* 2012; 27(2): 490–499.
- Serdio F, Lughofer E and Pichler K. Residual based fault detection using soft computing techniques for condition monitoring at rolling mills. *J Inform Sci* 2014; 259: 304–320.
- Gyftakis N, Spyropoulos V, Kappatou C, et al. A novel approach for broken bar fault diagnosis in induction motors through torque monitoring. *IEEE T Energy Convers* 2013; 28: 267–277.
- Murphey L, Masrur A, Zhihang C, et al. Model-based fault diagnosis in electric drives using machine learning. *IEEE/ASME T Mech* 2006; 11(3): 290–303.
- Lemos A, Caminhas W and Gomide F. Adaptive fault detection and diagnosis using an evolving fuzzy classifier. *J Inform Sci* 2013; 220: 64–85.
- Zheng Y, Fang H and Wang H. Takagi-Sugeno fuzzy-model-based fault detection for networked control systems with Markov delays. *IEEE T Syst Man Cy B* 2006; 36(4): 924–929.
- Wen X. *A hybrid intelligent technique for induction motor condition monitoring*. PhD Thesis, Institute of Industrial Research, University of Portsmouth, Portsmouth, 2011.
- Ding X. *Model based fault diagnosis techniques: design schemes algorithms and tools*. 1st ed. Berlin: Springer Verlag, 2013.
- Lughofer E. *Evolving fuzzy system methodologies, advanced concepts and applications, studies in fuzziness and soft computing*, vol. 266. Berlin, Heidelberg: Springer, 2011.
- Lughofer E. On-line assurance of interpretability criteria in evolving fuzzy systems achievements. *J Inform Sci* 2013; 251: 22–46.
- Gacto J, Alcalá R and Herrera F. Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures. *Inform Sciences* 2011; 181(20): 4340–4436.
- Magne S. Simplification and reduction of fuzzy rules. In: *Studies in fuzziness and soft computing modelling*, vol. 128. Springer Berlin Heidelberg, 2003, pp. 278–302.
- Lughofer E, Jean B and Ammer S. On-line elimination of local redundancies in evolving fuzzy systems. *Evol Syst* 2011; 2(3): 165–187.
- Lughofer E and Guardiola C. On-line fault detection with data-driven evolving fuzzy models. *Contr Intell Syst* 2008; 36(4): 307–317.
- Ahmadi H, Sakhaii B and Labbafi R. Support vector machine and K-nearest neighbour for unbalanced fault detection. *J Qual Mainten Eng* 2014; 20(1): 65–75.
- Xuhong W and Yigang H. Diagonal recurrent neural network based on-line stator winding turn fault detection for induction motors. In: *Proceedings of the IEEE eighth international conference on electrical machines and systems*, Nanjing, 2005, pp. 2266–2269.
- Serdio F, Lughofer E and Pichler K. Fault detection in multi-sensor networks based on multivariate time-series models and orthogonal transformation. *Inform Fusion* 2014; 20: 272–291.

18. Seyed M, Abdesslem D and Davod K. Fault detection in 3-phase Traction Motor Using Artificial Neural Networks. In: *IEEE conference on transportation electrification*, Dearborn, 2012, pp.1–6.
19. Widodo A and Yang S. Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors. *Expert Syst Appl* 2007; 33(1): 241–250.
20. Chiang H, Russell L and Braatz D. *Fault detection and diagnosis in industrial systems* (Advanced textbooks in control and signal processing). 1st ed. London: Springer, 2001.
21. Fuente, J. Fault detection in a wastewater treatment plant based on neural networks and PCA, In: *20th Mediterranean Conference on Control & Automation*, Barcelona, 3–6, July 2012, pp. 758–763.
22. Odgaard F, Lin B and Jorgensen S. Observer and data-driven-model-based fault detection in power plant coal mills. *IEEE T Energy Conver* 2008; 23(2): 659–668.
23. Khushaba N, Al-Ani A and Al-Jumaily A. Orthogonal fuzzy neighborhood discriminant analysis for multifunction myoelectric hand control. *IEEE T Bio-Med Eng* 2010; 57(6): 1410–1419.
24. Pichler K, Pichler M, Lughofer E, et al. On the robustness of fault detection in reciprocating compressor valves. In: *proceeding of the IEEE International Conference on Systems, Man and Cybernetics*, San Diego, 5–8 October 2014, pp.2733–2738. New York: IEEE.
25. Mitoma T, Wang H and Chen P. Fault diagnosis and condition surveillance for plant rotating machinery using partially-linearized neural network. *Comput Ind Eng* 2008; 55(4): 783–794.
26. Goharrizi Y and Sepehri N. A wavelet-based approach to internal seal damage diagnosis in hydraulic actuators. *IEEE T Ind Electron* 2010; 57(5): 1755–1763.
27. Seshadrinath J, Singh B and Panigrahi BK. Investigation of vibration signatures for multiple fault diagnosis in variable frequency drives using complex wavelets. *IEEE T Power Electr* 2014; 29(2): 936–945.
28. Yan R, Gao RX and Chen X. Wavelets for fault diagnosis of rotary machines: a review with applications. *Signal Process* 2014; 96: 1–15.
29. Antonino J, Aviyente S, Strangas G, et al. Scale invariant feature extraction algorithm for the automatic diagnosis of rotor asymmetries in induction motors. *IEEE T Ind Inform* 2013; 9(1): 100–108.
30. Antonino J, Riera M, Roger J, et al. Application and optimization of the discrete wavelet transform for the detection of broken rotor bars in induction machines. *Appl Comput Harmon A* 2006; 21(2): 268–279.
31. Li W, Monti A and Ponci F. Fault detection and classification in medium voltage DC shipboard power systems with wavelets and artificial neural networks. *IEEE T Instrum Meas* 2014; 63(11): 2651–2665.
32. Alok S, Kuldip K and Godfrey C. Class-dependent PCA, MDC and LDA: a combined classifier for pattern classification. *Pattern Recogn* 2006; 39(7): 1215–1229.
33. Prieto D, Cirrincione G, Espinosa G, et al. Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. *IEEE T Ind Electron* 2013; 6(8): 3398–3407.
34. Delgado M, Cirrincione G and Garcia Espinosa A. Bearing faults detection by a novel condition monitoring scheme based on statistical-time features and neural networks. *IEEE T Ind Electron* 2013; 60(8): 3398–3407.
35. Jolliffe T. *Principal component analysis*. 2nd ed. Berlin, Heidelberg; New York: Springer, 2002.
36. Eleyan A and Demirel H. PCA and LDA based neural networks for human face recognition. INTECH, 2007, pp. 93–106, Available from: http://www.intechopen.com/books/face_recognition/pca_and_lda_based_neural_networks_for_human_face_recognition.