PAN POLSKA AKADEMIA NAUK

METROLOGY AND MEASUREMENT SYSTEMS

Index 330930, ISSN 0860-8229

www.metrology.pg.gda.pl

M&MS

A NOVELTY DETECTION APPROACH TO MONITORING OF EPICYCLIC GEARBOX HEALTH

Ziemowit Dworakowski, Kajetan Dziedziech, Adam Jabłoński

AGH University of Science and Technology, Al. A. Mickiewicza 30, 30-059, Cracow, Poland
(✉ *zdw@agh.edu.pl*, 48 12 617 3396, *dziedzie@agh.edu.pl*, *ajab@agh.edu.pl*)

Abstract

Reliable monitoring for detection of damage in epicyclic gearboxes is a serious concern for all industries in which these gearboxes operate in a harsh environment and in variable operational conditions. In this paper, autonomous multidimensional novelty detection algorithms are used to estimate the gearbox' health state based on vectors of features calculated from the vibration signal. The authors examine various feature vectors, various sources of data and many different damage scenarios in order to compare novel detection algorithms based on three different principles of operation: a distance in the feature space, a probability distribution, and an ANN (*artificial neural network*)-based model reconstruction approach. In order to compensate for non-deterministic results of training of neural networks, which may lead to different network performance, the ensemble technique is used to combine responses from several networks. The methods are tested in a series of practical experiments involving implanting a damage in industrial epicyclic gearboxes, and acquisition of data at variable speed conditions.

Keywords: epicyclic gearbox, soft computing, auto-associative neural network, novelty detection, vibration signal.

1. Introduction

1.1. Rotary machinery monitoring

Structural health monitoring plays a vital role in all industrial areas as it enables both cost-efficient maintenance and avoidance of catastrophic failures. One of the general categories for structural health monitoring is focused on monitoring of rotary machinery (turbines, shafts, gearboxes, generators, *etc*.). Probably the most popular approach involves acquisition of vibration signatures, extraction of damage-sensitive features from these signals, and processing such data in order to determine whether or not any faults are present in the system [1–5].

There are three major approaches to this problem regarding data necessary for the decision process. Firstly, one can use only data acquired for an unknown state of the system and compare selected features (*e.g.* RMS) of the acquired signals with an appropriate norm (see Fig. 1a). In this way no information on the previous condition of a particular machine and no historical data are required. However, since all of the machines must meet similar requirements, the approach is very general, and fails in detecting small faults that do not cause a significant alteration of signal features. Secondly, one could gather a database of signals, including examples acquired

Article history: received on Oct. 19, 2017; accepted on Apr. 21, 2018; available online on Sep. 30, 2018, DOI: 10.24425/123896.

for similar machines in healthy and damaged states, which enable to design a supervised data classifier [6]. Signals similar to historically gathered *healthy data* would be classified as OK, while those similar to historically gathered *faulty data* would be classified as DAMAGE (see Fig. 1b). Unfortunately, such an approach is usually unfeasible because acquisition of signals representing multiple instances of various faults is too costly to be performed in practice. Tertiary, one could acquire signals for one particular machine monitoring its operation over a long period and recording the variability of signals in the healthy state. Then, the signals acquired in an unknown state can be compared with those collected in the database of *normal* signals – any significant differences serve as a premise of damage (see Fig. 1c). The third approach is known as *novelty detection*, *anomaly detection* or *outlier analysis* [7].

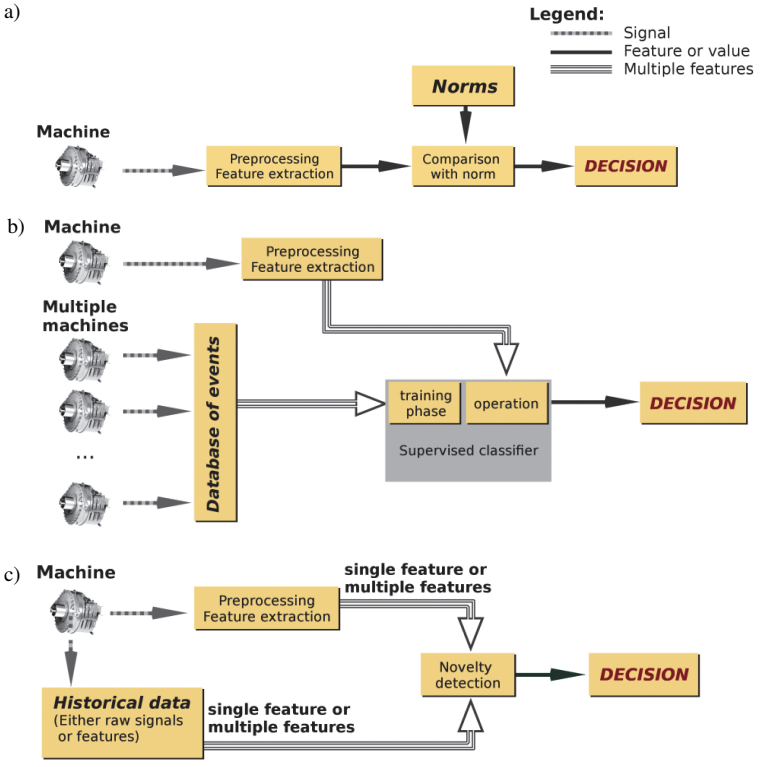

Fig. 1. Approaches to rotary machinery diagnostics regarding necessary data: Norm-based diagnose (a); supervised classification of faults (b); and trend analysis/novelty detection (c).

1.2. Novelty detection for monitoring of mechanical systems

In general, *Novelty Detection* (ND) is a task of recognizing data that are significantly different than data belonging to a *normal* class; thus, it is a one-class-classification problem. Each sample either is a member of a *normal* class, or is not (and should be marked as *novel*). The approach is very popular in monitoring of mechanical systems and has previously been used to diagnose bridges [8], plate-like structures [9], aircrafts [10], and many more [11–14].

In vibro-diagnostics, a simple ND-based unidimensional approach called *trend analysis* has become a standard feature in all condition-monitoring systems: values of separate features (*e.g.* RMS or Peak-to-peak) are plotted against time. In this scenario, variability of features in *normal*

conditions is used to set up a threshold for damage detection. The multidimensional novelty detection usually requires calculation of a distance between the *normal* and newly acquired data. Different metrics can be used to this end, *e.g.* Euclidean or Mahalanobis distance, which were used for bearing fault detection [15–17] or planetary gearbox diagnosis [18]. More advanced solutions require calculation of a boundary around the *normal* region. Here, a neural-network-based or support-vector-machine-based approaches are probably the most common.

Examples related to vibro-diagnostics include *e.g.* [19–20]. PCA-based *confidence region* models were used for detection of tooth defects in helical gears [21] or bearings [17].

Another approach to novelty detection requires computation of a model that, given the input to the object under monitoring, enables prediction of the output. Faults are detected by comparison of the model-obtained output with the measured one. Features calculated from such a residual signal may serve as damage indicators. Examples of methods for this approach include most notably autoregressive models [22–23] or *auto-associative neural networks* (AANNs) which, although not yet implemented in vibro-diagnostics, were used *e.g.* for detection of damage based on frequency-response functions [24] or guided-wave-based signals [25].

Novelty detection enables monitoring of rotary machinery in nonstationary operational and environmental conditions, provided that all significant environmental and operational states are well represented in the training data.

1.3. Contribution and organization of this paper

The paper is dedicated to verification and comparison of efficiency of four different novelty detection methods in the task of epicyclic gearbox health assessment. The authors implemented a *nearest-neighbour* (NN) distance-based method, a *data-distribution-based* (DDB) method and two model-based solutions involving AANNs. All the methods take multidimensional feature vectors as inputs. To the best of the authors' knowledge, all the three methods are novel in terms of their application, *i.e.* they have never been applied to detection of damage in epicyclic gearboxes, although the DDB and NN methods were previously used in vibro-diagnostics [17]. Moreover, the ensemble AANN method provides novelty by itself as ensembles of AANNs have never been tested in any vibro-diagnostic-based novelty detection task.

As claimed by Shen [26], in terms of fault identification, standard techniques are limited to planetary gear because of changing transmission path, moving fault location, and modulations. In order to tackle this problem, the test rig has been equipped with epicyclic gearboxes characterized by relatively large ratios. These ratios, resulting in a relatively slow speed, emphasize the problem of complexity and low energy of fault-generated components.

The remainder of the paper is organized as follows: Section 2 provides a description of methods used for detection of novelty in this work and introduces the features extracted from vibrational signals, Section 3 provides a description of experimental procedures, test bench, organization of feature vectors, sensors, description of damages, and results obtained with all the methods. Finally, Section 4 summarizes and concludes the paper.

2. Applied methods

2.1. Novelty detection algorithms

2.1.1. Nearest-neighbour approach

A simple nearest-neighbour approach consists in calculation of the distance between all training data points and a newly acquired data sample. The minimum distance is compared with

a threshold: Distances exceeding the threshold are treated as a premise of *novelty* of the sample. The threshold is calculated based on the average minimum distance between data samples in the training dataset + three times the standard deviation of this value. The method is conceptually simple, although requiring a significant computational effort, as each new sample is compared with each point in the dataset. The method is also vulnerable to outliers in the training data. A graphical illustration of the method is presented in Fig. 2.

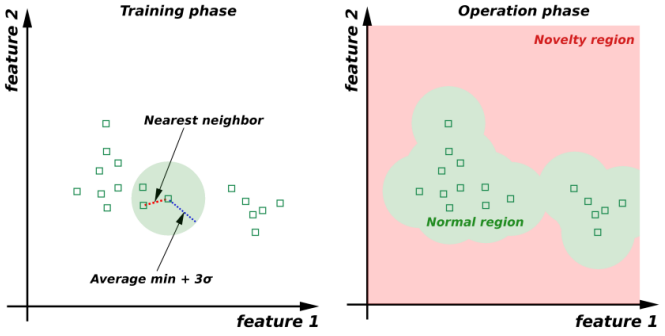

Fig. 2. The principle of operation of a nearest-neighbour-based novelty detection.

2.1.2. Unidimensional distribution-based approach

A standard unidimensional trend analysis concept can easily be extended to multiple dimensions by dealing with each dimension separately. For each dimension the probability distribution should be calculated. Newly acquired samples are compared against this distribution – if they fall outside of the plausible region they are marked as novel ones. The simplest approach to solving this in practice is to assume that the distribution is normal (note that this assumption can be misleading!) and calculate the average x value and standard deviation σ of samples. Then, two thresholds denoting the range of normal data can be defined as $x-3\sigma$ and $x+3\sigma$, where x is a mean of samples. In the case when the abovementioned assumption is not met (*i.e.* the data do not have a normal distribution), the method may produce wrong results, *e.g.* for a bimodal distribution it tends to be overconfident, marking many *novel* samples as *normal* ones – see Fig. 3.

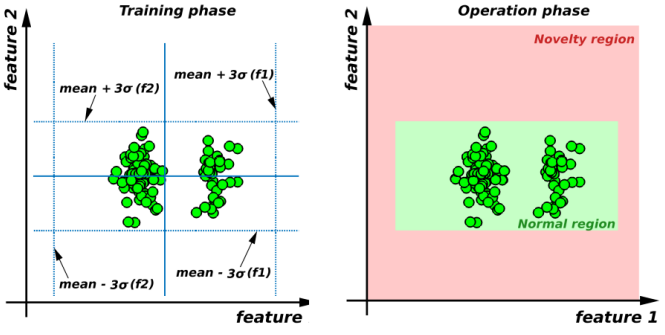

Fig. 3. The principle of operation of a simple distribution-based novelty detection. Although the method works good for dimensions in which data have a normal distribution (see thresholds for the normal region for feature 2, it tends to produce many false negatives otherwise (see thresholds for the normal region for feature 1.

2.1.3. Auto-associative neural network

Auto-associative neural networks are essentially *multi-layered perceptrons* (MLPs) [6, 27–28] that are trained by presenting to the network target vectors equal to the training ones, *i.e.* the network is expected to produce the input at its output. The network is designed using a *bottleneck* approach, that is: one of hidden layers have a small number of neurons, which forces the network to develop generalized representations of the training data (see Fig. 4a). In the operational phase such a network is presented with new patterns. The output of the network is compared with the presented pattern. The distance between input and output is treated as the extent of novelty of a data sample: samples for which distances are higher than the median distance $+3\sigma$ are denoted as *novel*. For determination of AANN's scale well into higher dimensions (*i.e.* they are not as affected by dimensionality of the problem as some other solutions) however, their results are non-deterministic due to a random nature of their training: the same network trained on relatively similar data might learn significantly different rules for separating normal and abnormal regions in the data space (see Fig. 4b). For that reason it is a good approach to combine multiple neural nets in a so-called *ensemble approach* – in which responses of many individual AANNs are combined in order to increase repeatability of results.

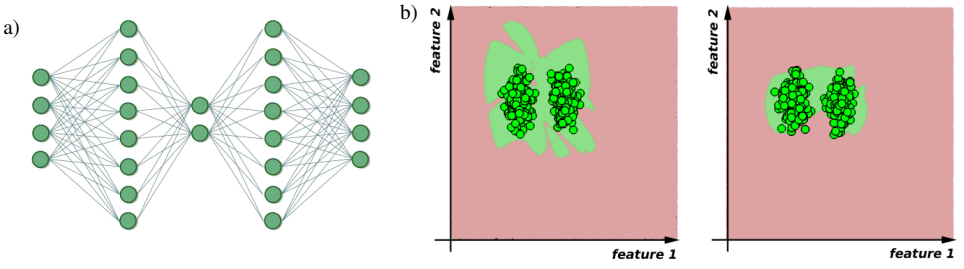

Fig. 4. A scheme of the AANN network (a) and data classification rules produced by two similar networks trained on data from a similar distribution (b).

In this paper, AANN consisted of three hidden layers with 10, 2 and 10 neurons, respectively. The input layer size was varying depending on the input vector size (see Table 4). The output layer size was equal to that of the input. The network was trained with the Levenberg-Marquardt backpropagation algorithm. Since the paper evaluates the influence of input vector, source of data, various algorithms and many damage cases, the optimization of the network structure was not performed rigorously, as it would extend the required number of tests by the order of magnitude and would cause the paper to be unclear. Instead, the size of networks was based on the authors' experience and preliminary examinations. It was not found that the network size has a large influence on the final results, provided that the bottleneck is preserved and the network is not too large to prevent overfitting. The influence of the initial weights' distribution is far more important.

2.2. Signal features

Many different damage-sensitive features of vibrational signals have been proposed in the past decades, ranging from simple signal statistics to more advanced high-level features aimed at detection of particular faults. The most frequently used simple ones are *root mean square* (RMS), crest factor, peak-to-peak and kurtosis of signals (*e.g.* [29–31]). Some more advanced have also been reviewed by different researchers [1, 32]. All these metrics can be calculated either for

a raw signal or for a pre-processed one. In this work several simple features are used as inputs for novelty detection algorithms. A brief description of them is provided in Table 1.

Table 1. Signal features used as inputs for the algorithms. f_c refers to the characteristic frequency of an input.

Feature number	Feature name	Description
0	Speed	Rough speed estimate (constant for the whole sigal)
1	PP	Peak-to-peak value of the signal
2	RMS	Root-mean-square of the signal
3	Kurtosis	Kurtosis (fourth central moment) of the signal
4	Crest	Crest factor (PP/RMS) of the signal
5	Skew	Skewness (Third central moment) of the signal
6	SpecPP	PP value of the spectrum
7	SpecKurt	Fourth central moment of the spectrum
8	SpecSkew	Third central moment of the spectrum
9	LowSpecRMS	RMS of a low-frequency part of the spectrum: frequency range from 0 to $0.5f_c$
10	MedSpecRMS	RMS of a medium-frequency part of the spectrum: frequency range from $0.5f_c$ to $2f_c$
11	HiSpecRMS	RMS of a high-frequency part of the spectrum: frequency range from $2f_c$ to $10f_c$
12	VHiSpecRMS	RMS of a very high-frequency part of the spectrum: frequency range above $10f_c$

3. Experimental evaluation

3.1. Test bench description

Experimental tests have been performed on a modified AMC Vibro VibStand 2 test bench. The VibStand 2 test bench includes a driving electric motor followed by a parallel gearbox with a reduction ratio 2.91:1 The modification included adding two 2-stage planetary gearboxes NEUGART PLQE060-064-SSSA3AD-R10 of a ratio 64:1 connected in series after the parallel shaft gearbox. The second planetary gearbox was acting as the loading for the first planetary gearbox. The test bench is presented in Fig. 5. A kinematic scheme of a gearbox is given in Figure 6. Data

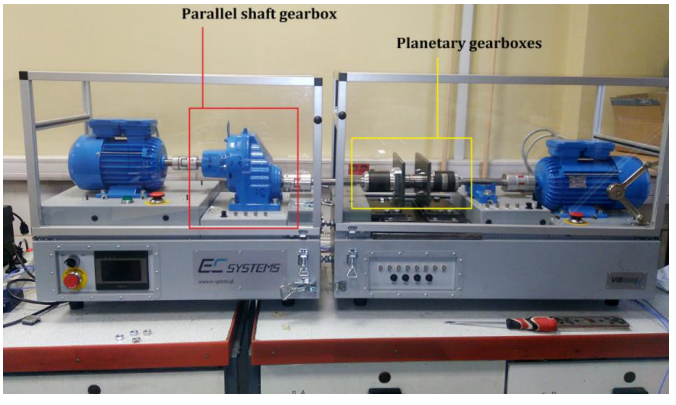

Fig. 5. The test bench used in experiments.

were acquired with an AMC Vibro VibMonitor. A sampling frequency was 25 kHz and an acquisition time was 1 s for each signal, which gave a total of 25 600 samples. Data were gathered with three different sensors, *i.e.* a Polytec PSV-400 Doppler Laser Vibrometer measuring velocity; a PCB 333B30 piezoelectric accelerometer; and a PCB 740B02 piezoelectric strain sensor. The sensors were measuring responses on the first planetary gearbox. Additionally, a reference speed was measured directly from the AMC Vibro VibStand 2 test bench on the driving motor with an encoder.

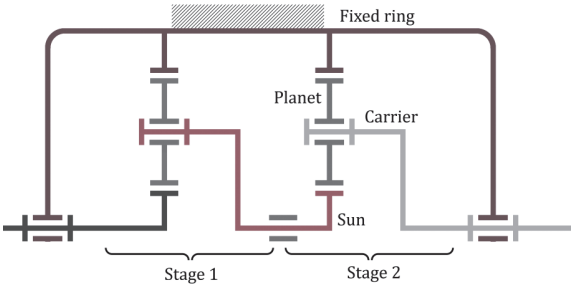

Fig. 6. A kinematic scheme of a single planetary gearbox.

3.2. Signal acquisition routines

The operational data were acquired in 6 sessions. The first session represented a healthy gearbox. Every consecutive session represented a different *health state* (HS) of the monitored gearbox. In these sessions damages were implanted in the gear train as a degradation of one tooth. Once a tooth was removed, this damage was present in all of the following sessions. All health states are described in Table 2 and pictured in Fig. 7. The signals for all health states are presented in Fig. 8. Between sessions, the test bench was reassembled (all gearboxes and sensors were reattached). In every session data were measured at constant speeds at a number of different levels, *i.e.* speed was measured in a range of 1000–3000 rpm with a step of 100 rpm, so in total 21 levels of different speeds were measured. For each level of speed one 300-second-long signal was acquired and later divided into 1-second-long signal parts for the purpose of data analysis. For each of these 1-second-long signal parts the features were calculated separately.

Table 2. Gearbox health states examined during experiments.

HS number	Description
1	Healthy
2	Removed tooth from ring gear related to the first stage of the PG (see Fig. 7a)
3	Above + removed tooth from the sun gear related to the second stage of the PG (see Fig. 7b)
4	Above + removed tooth from the sun gear related to the first stage of the PG (see Fig. 7c)
5	Above + removed tooth from the planet gear related to the second stage of the PG (see Fig. 7d)
6	Above + removed tooth from the planet gear related to the first stage of the PG (see Fig. 7e)

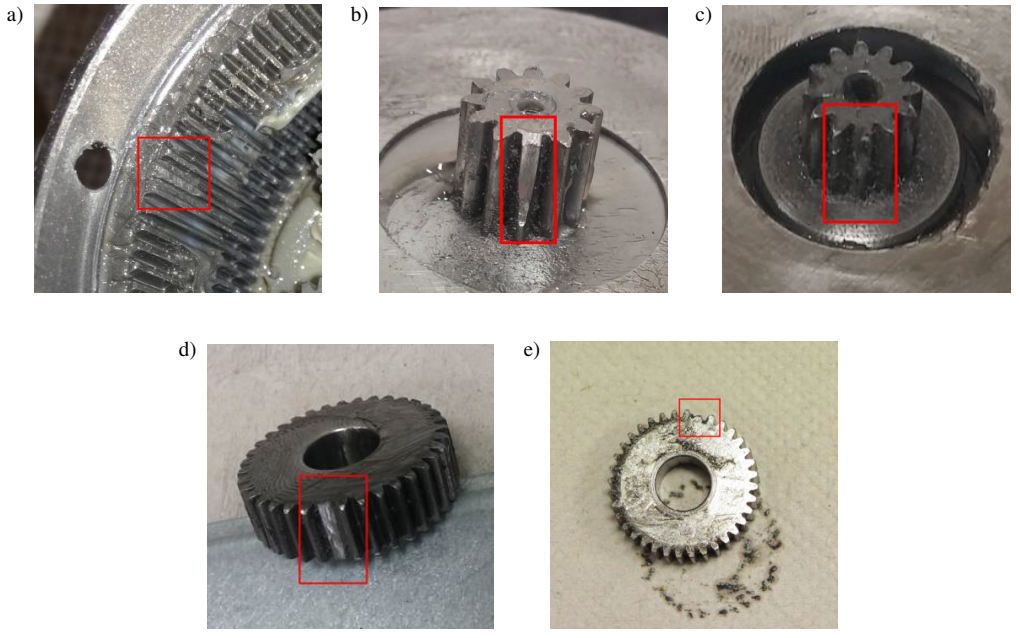

Fig. 7. Damages implanted in the gearbox. A tooth removed from the ring gear related to the first stage of the PG (a); a tooth removed from the sun gear related to the second stage of the PG (b); a tooth removed from the sun gear related to the first stage of the PG (c); a tooth removed from one of the planet gears related to the second stage of the PG (d); a tooth removed from one of the planet gears related to the first stage of the PG (e).

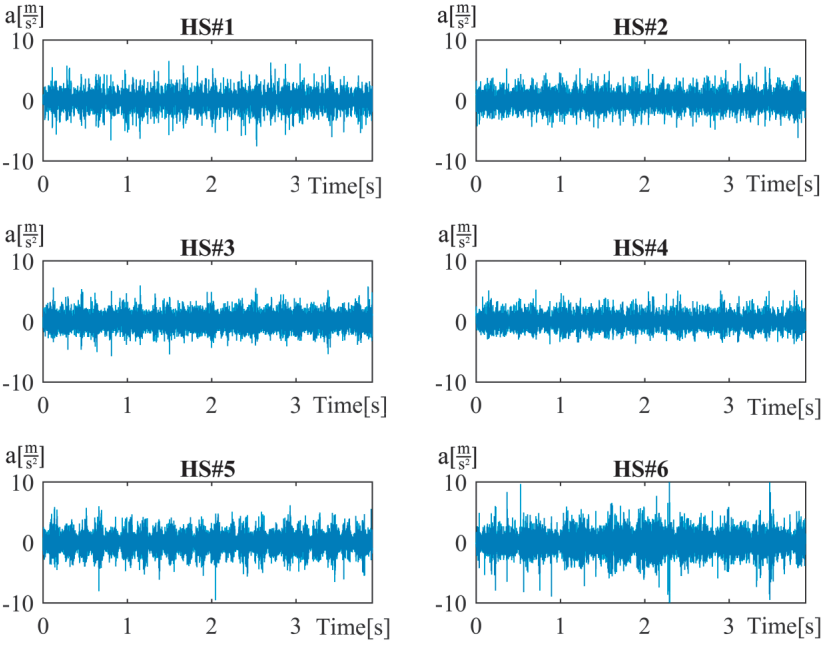

Fig. 8. Fragments of vibration signals acquired for 1000 rpm for different health states.

3.3. Methods' training

Acquired 1-second-long signal parts were gathered using three different sensors, 13 different features were calculated from each signal, 6 different cases of damage were present during tests. In order to fully verify the contribution of various features, the number of features used, various sensor configurations and various cases of damage, each method test was assigned with a particular feature vector, a particular sensor or a set of sensors and a particular arrangement of the reference and testing data. Due to a great number of possible feature combinations (2^{13}), only arbitrarily selected feature vectors were examined. The features are defined in Table 1.

In Table 3 these features are distributed over various possible vectors used as inputs. All features were normalized in respect of the intact data. In all the cases the data registered at 21 different speeds were used together. In Table 4 various configurations of input sensors are presented. Note, that the number of input features is the product of the feature vector size and the number of input sensors. Finally, Table 5 contains all configurations of the training and testing

Table 3. Feature configurations of input feature vectors.

Feature vector number	Features used
1	PP, RMS, Kurtosis
2	Crest, Skew, SpecPeak
3	SpecKurt, SpecSkew, LowSpecRMS
4	MedSpecRMS, HiSpecRMS, VHiSpecRMS
5	Speed + PP, RMS, Kurtosis
6	Speed + Crest, Skew, SpecPeak
7	Speed + SpecKurt, SpecSkew, LowSpecRMS
8	RMS, Skew, MedSpecRMS, HiSpecRMS, VHiSpecRMS
9	Kurtosis, Crest, SpecKurt, SpecSkew, LowSpecRMS
10	All features with speed
11	All features without speed

Table 4. Input sensors used.

Sensor configuration	1	2	3	4
Sensors used	Vibrometer	Strain sensor	Accelerometer	All sensors

Table 5. Reference and test data sources.

Data arrangement	Source of training data	Source of testing data
1	HS#1 (intact)	HS#2
2	HS#1 (intact)	HS#3
3	HS#1 (intact)	HS#4
4	HS#1 (intact)	HS#5
5	HS#1 (intact)	HS#6
6	HS#2	HS#3
7	HS#3	HS#4
8	HS#4	HS#5
9	HS#5	HS#6

data. For each damaged case either the initial (intact) state or the previous damage state were used as a reference.

For the purpose of training, the reference data were divided randomly, so that 80% of data would be used for training of the novelty detectors, while the remaining 20% would serve for calculation of the false positive rate (*i.e.* how often a *normal* state triggers detection of a novelty).

3.4. Results of monitoring phase

3.4.1. Novelty detection efficiency

None of the examined methods was able to detect a great number of novelties for all configurations of inputs. Some input configurations, however, enabled the methods to obtain very good results. In particular, the methods that used as features RMS of selected frequency bands of a spectrum enabled to detect almost 100% novelties for data arrangement #5.

In general, it is evident that the NN method is a more reliable approach than the two other examined methods. Even if the "best" features are only a part of the input vector, the method still gives consistently good results (see *e.g.* Fig. 9d): feature vectors #4, #8, #10 and #11 all include features #10, #11 and #12 which seem to be suited particularly well for this task. AANN, on the other hand, provides a similar efficiency for feature vectors #4 and #8 but is unable to work well if there is a large number of inputs besides the *good ones*, as in the case of feature vectors #10 and #11 (see Fig. 9f).

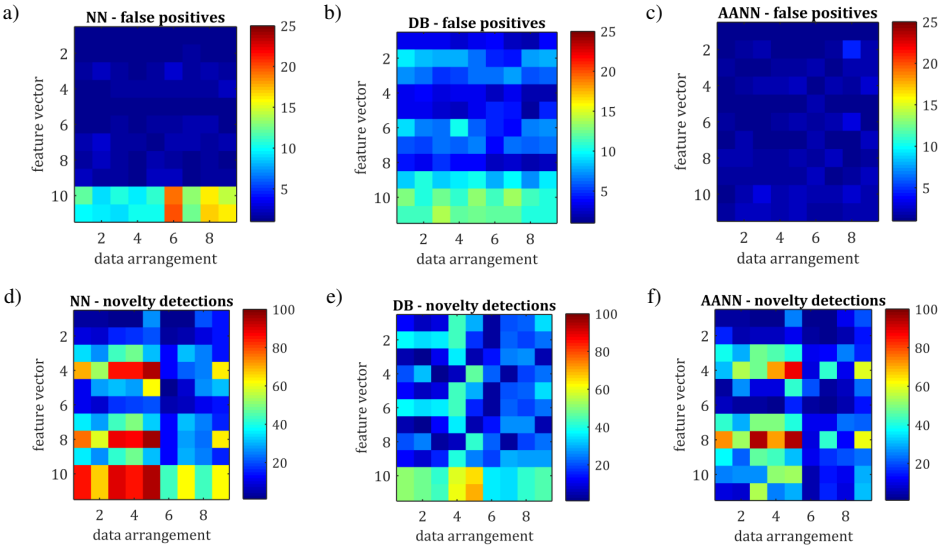

Fig. 9. Percentage of false positives (a–c) and efficiency of novelty detection (true positives) (d–f) of the proposed methods for different arrangements of data and different feature vectors used. For all results provided in this figure sensor configuration #4 was used as the input (all the sensors). The colour version of these figures the reader can find in the web version of this paper. NN, false positive rate (a); DDB, false positive rate (b); AANN, false positive rate (c); NN, novelty detection (d); DDB, novelty detection (e); AANN, novelty detection (f).

Out of the three examined sensors, the accelerometer provided signals of the highest potential for damage detection – which can easily be seen in Figs. 10e and 10f. Again, the requirement for good performance of NN was that the accelerometer would only be included among others

to perform well. On the other hand, for comparative performance of AANN it was required that only the data from the accelerometer were used as the input – in Fig. 10f the efficiency for the 3rd sensor configuration is significantly better than that for the 4th.

The DDB method appears to be less vulnerable to feature choice with a similar but relatively lower efficiency for all examined feature vectors. Contrary to the NN and AANN, the best results were obtained for large feature vectors – which is not surprising as a novelty is detected for each dimension separately: the greater the number of dimensions, the higher is also the chance to detect a novelty in data (see Fig. 9e). Consistently, the method is also better for novelty detection if more sensors are used as inputs (see Fig. 10b). It is worth noticing, however, that the number of false positives also rises in this case (see Fig. 10e).

3.4.2. False positives

In general, NN-based and AANN-based approaches consistently provide a small number of false positive indications – the value of error ranges from 0 to roughly 3 per cent, with the exception of a very large number of input features for the NN approach. In Fig. 9a two bottom rows of results show that the method resulted in up to 20% of false positives. These rows correspond to the input feature numbers 36 and 33, respectively. The results are consistent with those presented in Fig. 10a. The number of false positives is small for feature vectors #1 – #9 or single sensors used as outputs. However, if all the three sensors and all features are included, the false positive rate is significantly higher.

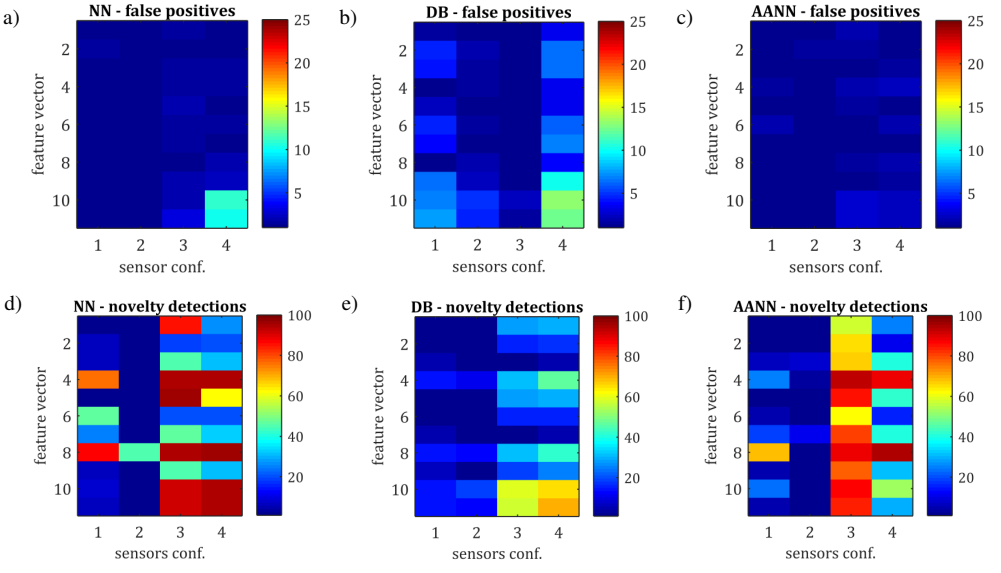

Fig. 10. Percentage of false positives (a–c) and true positives (d–f) of the proposed methods for different sensors used as inputs and different feature vectors used. For all results provided in this figure data arrangement #5 was used (the intact state as a reference, all damages as a *novelty*). The colour version of these figures the reader can find in the web version of this paper. NN, false positive rate (a); DDB, false positive rate (b); AANN, false positive rate (c); NN, novelty detection (d); DDB, novelty detection (e); AANN, novelty detection (f).

The false positive rate for AANN method does not seem to be vulnerable to sensor, feature or damage combinations and remained very low. In contrast, the DDB method resulted in great numbers of false positive indications. Although for some feature vectors this phenomenon was

not observed (*e.g.* feature vectors #1, #4, #5 and #8), in the remaining cases the false positive rate was relatively high (5–13%).

3.4.3. Best case scenario

The above analyses show that the feature selection and choice of a data source are essential for design of any novelty detection system. In the examined case, it was decided that feature vector #8 calculated for signals acquired with the accelerometer should be the preferred choice. For such a setup, the methods were examined once more in detection of all damage cases. Each method was tested 10 times, each time with a random division of reference data into the training and testing subsets. Due to a high variability of results for AANN, responses of five nets trained on the same subsets of data were used to form an ensemble: the majority voting was used to determine whether a given sample belongs to a *normal* or *novel* class. Due to poor performance in previous tests, the DDB classifier was not included in this evaluation.

In Fig. 11 the results of such consecutive runs of the simulation for feature vector #8, sensor #3 and all data configurations are provided. It is clear that repeatability of results for AANN network is very low. Although usually AANNs produce good results, at times training converges to a point with a very low testing efficiency, either in terms of high false positive rate or low novelty detection efficiency. For instance, for data arrangement #3 (3rd case of damage with the healthy state as a reference) the vertical line denotes that the spread of results covers almost full range from 0 to 100%. Even though the majority of results have well above 90% efficiency, first 25% of results have their efficiency spreading from 0 to 90%, meaning that few ANNs exist, which failed in detecting all the novelties.

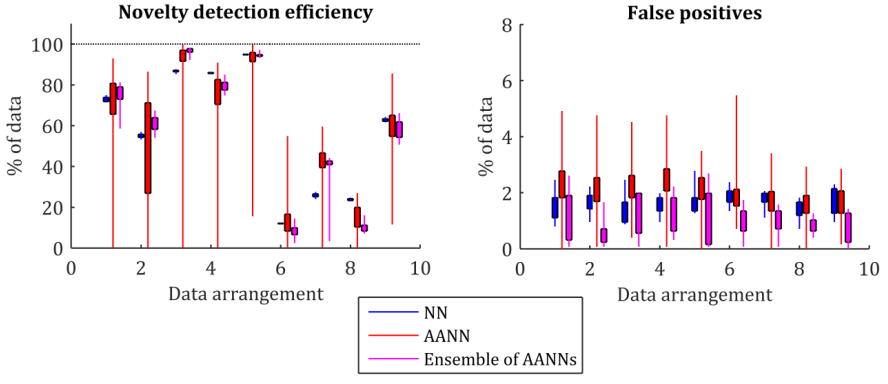

Fig. 11. Repeatability of results of the three algorithms, namely: nearest-neighbour approach, AANN approach and ensemble-of-AANNs approach. Peak-to-peak spreads of results are marked with vertical lines. A range from Q_1 to Q_3 is marked with rectangles. The colour version of these figures the reader can find in the web version of this paper.

On the other hand, using even a very simple ensemble of AANNs, consisting of five networks only and without any advanced ensemble design technique, significantly boosts performance of neural nets to the point of outperforming the nearest-neighbour-based approach. Although the spread of results for the ensemble approach is still notable (probably it could have been lowered further using a more advanced ensemble design and bigger ensembles) it outperforms the NN solution for all data arrangements in terms of false positive rate and is better than NN for data arrangements #1, #2, #3 and #7 in terms of novelty detection efficiency.

4. Summary and conclusions

In this work four different novelty detection methods for epicyclic gearbox health monitoring were implemented and tested. The inputs to the algorithms consisted of various configurations of 12 features calculated from vibration signals and three different sensors used as sources of data. The features calculated from selected bands of the frequency spectrum appeared to be the most suitable for the task at hand. The accelerometer provided signals of a better quality for the state change detection than the laser vibrometer and the strain sensor. It was found that the ensemble approach based on auto-associative neural nets provides the smallest number of false positive indications while maintaining efficiency comparable to the nearest-neighbour approach. Since implanting a damage required reassembly of the gearbox, the detected *novelties* might be a result of reassembly itself, not the damage-related changes in signals. Nevertheless, the method is able to efficiently detect all examined state changes in a gearbox operating at input speeds in a range from 1000 to 3000 RPM. The variability of results obtained with neural nets is expected to be lower if more ANNs are used to form an ensemble.

Acknowledgement

The work presented in this paper was supported by the National Centre for Research and Development in Poland under the research project no. PBS3/B6/21/2015.

References

[1] Samuel, P.D., Pines, D.J. (2005). A review of vibration-based techniques for helicopter transmission diagnostics. *J. Sound Vib.*, 282(1–2), 475–508.

[2] Kandukuri, S.T., Klausen, A., Karimi, H.R., Robbersmyr, K.G., (2016). A review of diagnostics and prognostics of low-speed machinery towards wind turbine farm-level health management. *Renew. Sustain. Energy Rev.*, 53, 697–708.

[3] Jardine, A.K.S., Lin, D., Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.*, 20, 1483–1510.

[4] Zimroz, R., Urbanek, J., Barszcz, T., Bartelmus, W., Millios, F., Martin, N. (2011). Measurement of instantaneous shaft speed by advanced vibration signal processing – Application to wind turbine gearbox. *Metrol. Meas. Syst.*, 18(4), 701–712.

[5] Urbanek, J., Barszcz, T., Sawalhi, N., Randall, R.B. (2011). Comparison of amplitude-based and phase-based methods for speed tracking in application to wind turbines. *Metrol. Meas. Syst.*, 8(2), 295–304.

[6] Worden, K., Staszewski, W.J., Hensman, J.J. (2011). Natural computing for mechanical systems research: A tutorial overview. *Mech. Syst. Signal Process.*, 25(1), 4–111.

[7] Pimentel, M.A.F., Clifton, D.A., Clifton, L., Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, (99), 215–249.

[8] Omenzetter, P., Brownjohn, J.M.W., Moyo, P. (2004). Identification of unusual events in multi-channel bridge monitoring data. *Mech. Syst. Signal Process.*, 18(2), 409–430.

[9] Mustapha, F., Manson, G., Worden, K., Pierce, S.G. (2007). Damage location in an isotropic plate using a vector of novelty indices. *Mech. Syst. Signal Process.*, 21(4), 1885–1906.

[10] Papatheou, E., Manson, G., Barthorpe, R.J., Worden, K. (2014). The use of pseudo-faults for damage location in SHM: An experimental investigation on a Piper Tomahawk aircraft wing. *J. Sound Vib.*, 333(3), 971–990.

[11] Haggett, S.J., Chu, D.F. (2009). Evolving novelty detectors for specific applications. *Neurocomputing*, 72, 2392–2405.

[12] Rizzo, P., Sorrivi, E., Lanza di Scalea, F., Viola, E. (2007). Wavelet-based outlier analysis for guided wave structural monitoring: Application to multi-wire strands. *J. Sound Vib.*, 307(1–2), 52–68.

[13] Worden, K., Sohn, H., Farrar, C.R. (2002). Novelty Detection in a Changing Environment: Regression and Interpolation Approaches,. *J. Sound Vib.*, 258(4), 741–761.

[14] Worden, K., Manson, G., Fieller, N.R.J. (2000). Damage Detection Using Outlier Analysis. *J. Sound Vib.,* 229(3), 647–667.

[15] Klein, R. (2013). A Method for Anomaly Detection for Non-stationary Vibration Signatures. *Annu. Conf. Progn. Heal. Manag. Soc.*, 1–7.

[16] Tao, X., Lu, C., Wang, Z. (2013). An approach to performance assessment and fault diagnosis for rotating machinery equipment. *EURASIP J. Adv. Signal Process.*, 1, 1–8.

[17] Georgoulas, G., Loutas, T., Stylios, C.D., Kostopoulos, V. (2013). Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition. *Mech. Syst. Signal Process.*, 41(1–2), 510–525.

[18] Bartkowiak, A., Zimroz, R. (2011). Outliers analysis and one class classification approach for planetary gearbox diagnosis. *J. Phys. Conf. Ser.*, 305, 12031.

[19] Khawaja, T.S., Georgoulas, G., Vachtsevanos, G. (2008). An efficient Novelty Detector for online fault diagnosis based on Least Squares Support Vector Machines. *2008 Ieee Autotestcon*, 1, 1–6.

[20] Pirra, M., Fasana, A., Garibaldi, L., Marchesiello, S. (2012). Damage identification and external effects removal for roller bearing diagnostics. *European Conference of the Prognostics and Health Management Society*, 1–8.

[21] Baydar, N., Chen, Q., Ball, A., Kruger, U. (2001). Detection of Incipient Tooth Defect in Helical Gears Using Multivariate Statistics. *Mech. Syst. Signal Process.*, 15, 303–321.

[22] Yang, M., Makis, V. (2010). ARX model-based gearbox fault detection and localization under varying load conditions. *J. Sound Vib.*, 329(24), 5209–5221.

[23] Komorska, I. (2012). Automobile gearbox diagnostics on the basis of the reference model. *Mech. Control*, 31(1), 6–15.

[24] Dervilis, N., Choi, M., Antoniadou, I., Farinholt, K.M., Taylor, S.G., Barthorpe, R.J., Park, G., Worden, K., Farrar, C.R. (2012). Novelty detection applied to vibration data from a CX-100 wind turbine blade under fatigue loading. *J. Phys. Conf. Ser.*, 382, 12047.

[25] Nazarko, P., Ziemianski, L. (2016). Damage detection in aluminum and composite elements using neural networks for Lamb waves signal processing. *Eng. Fail. Anal.*, 69, 97–107.

[26] Sheng, S. (2012). Wind Turbine Gearbox Condition Monitoring Round Robin Study Vibration Analysis. *NREL/TP-5000-54530, Tech. Rep. July, NREL.*

[27] Coral, R., Flesch, C., Penz, C., Roisenberg, M., Pacheco, A. (2016). A monte carlo-based method for assessing the measurement uncertainty in the training and use of artificial neural networks. *Metrol. Meas. Syst.*, 23(2), 281–294.

[28] Dudzik, S. (2013). Characterization of material defects using active thermography and an artificial neural network. *Metrol. Meas. Syst.*, 20(3), 491–500.

[29] Wu, B., Saxena, A., Khawaja, T.S., Patrick, R., Vachtsevanos, G., Sparis, P. (2004). An Approach To Fault Diagnosis of Helicopter Planetary Gears. *IEEE Autotestcon*, 475–481.

[30] Li, R., He, D., Bechhoefer, E. (2009). Investigation on Fault Detection for Split Torque Gearbox Using Acoustic Emission and Vibration Signals. *Annual Conference of the Prognostics and Health Management Society*, 1–11.

[31] El-Morsy, M.S., Abouel-Seoud, S.A., Rabeih, E. (2010). Geared System Condition Diagnostics Via Torsional Vibration Measurement. *Proceedings of ISMA2010*, 2831–2842.

[32] Sharma, V., Parey, A. (2016). A Review of Gear Fault Diagnosis Using Various Condition Indicators. *Procedia Eng*., 144, 253–263.