Agricultural vehicle condition monitoring system based on unsupervised novelty detection

Dimitrios Moshou[1], Natsis Athanasios[2], Dimitrios Kateris[1], Ioannis Gravalos[3], Nader Sawalhi[4], Ioannis Kalimanis[2], Theodoros Gialamas[3], Panagiotis Xyradakis[3], Zisis Tsiropoulos[3]

[1]Department of Hydraulics, Soil Science and Agricultural Engineering, Aristotle University, School of Agriculture, P.O. 275, 54124, Thessaloniki, Greece, e-mail: dmoshou@agro.auth.gr
[2]Agricultural University of Athens, Department of Exploitation of Natural Resources and Agricultural Mechanics, Iera odos 75, 11855, Athens, Greece
[3]Technological Educational Institute of Larissa, School of Agricultural Technology, Department of Biosystems Engineering, 41110 Larissa, Greece
[4]School of Mechanical and Manufacturing Engineering, University of New South Wales, Sydney 2052, Australia

Abstract. A tractor gearbox test rig has been used to collect signals from different types of bearing faults. For vibration monitoring accelerometers have been used to obtain vibration data. For fuel-injectors a bearing checker has been used in order to collect acoustic data. One class Self Organizing Maps (OCSOM) are used for detecting faults when exposed to actual data from the system representing a yet unknown state. Feature extraction was performed using seven features. The feature vectors are then fed to the OCSOM for training. OCSOM classification gave promising results (more than 95% correct classification). The fusion of features from both the vertical and the horizontal accelerometer resulted in accurate fault detection. In the case of the fuel-injectors the feasibility of using one-class SOM has been tested in the detection of signal deviations indicating failure with high detection performance.

Keywords: novelty detection, condition monitoring, neural networks, unsupervised learning, self organizing map.

1 Introduction

The use of vibration signals is quite common in the field of condition monitoring of rotating machinery. By comparing the signals of a machine running in normal and faulty conditions, detection of faults like mass unbalance, rotor rub, shaft misalignment, gear failures and bearing defects is possible. These signals can also be used to detect the incipient failures of the machine components, through an on-line monitoring system, reducing the possibility of catastrophic damage and the down time. The procedure of fault diagnosis starts with data acquisition, followed by feature extraction, fault detection and identification. Feature extraction is critical for the success of the diagnostic procedure. Extended defects in the inner and outer races are common in rolling element bearings (see an example in Fig. 1).

The use of vibration signals is quite common in the field of condition monitoring and fault diagnosis of bearings (Xu et al., 2009). To inspect raw vibration signals, a

wide variety of techniques have been introduced that may be categorized into two main groups: classic signal processing (McFadden and Smith, 1984) and intelligent systems (Paya et al., 1997).

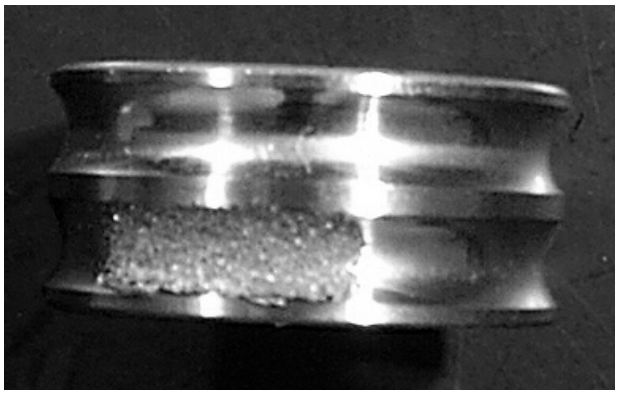

Fig. 1. Example of an extended fault in the inner race.

In the current work vibration monitoring is applied in the health condition monitoring and fault detection of two tractor components, the tractor gear box and the fuel-injectors. An approach from artificial intelligence, Self Organizing Maps (SOM) are used in the form of One Class SOM for detecting deviations in the vibration response of faulty bearings and subsequently in the acoustic response of fuel-injectors associated with malfunction due to wear.

2 Materials and Methods

Two experimental platforms (one for bearings and one for fuel-injectors) have been developed and used for commissioning experiments on fault detection and performing data acquisition. These data were further processed to extract specific features and develop novelty detection techniques relevant to fault presence. Details are presented in the following sections.

2.1 Gear box test platform data acquisition

A gearbox test rig has been used in order to collect signals from different types of bearing faults. A photograph of the rig showing the position of the accelerometers and the encoder at the output shaft is shown in Fig. 2 (Sawalhi, 2007). Two types of faults (inner race and outer race crack) were tested under a 50 Nm load, while setting the output shaft speed to 10 Hz (600 rpm). Vibration signals were collected using two accelerometers positioned on the top of the gearbox casing above the defective bearing (vertical accelerometer) and sideways respectively (horizontal

accelerometer). The 1.35 seconds (65536 samples) signals were sampled at 48 kHz. A photo-reflective switch was placed near the output shaft to measure its speed by providing a once per rev tacho signal. The torque for each case was measured at the input shaft.

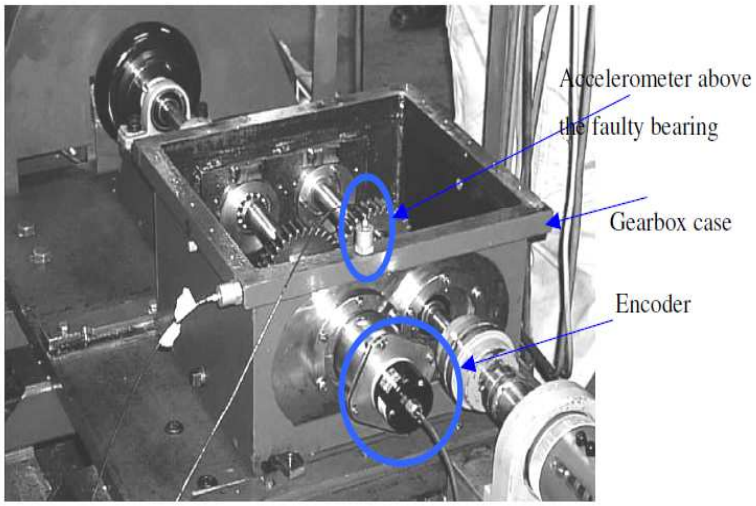

Fig. 2. The spur gear rig.

2.2 Fuel-injector data acquisition

The Bearing Checker (manufactured by SPM Instrument) was used for the fuel-injector measurements (Fig. 3). Normally, this instrument is used to measure the level of impulse during operation of the machine via an embedded microprocessor impulse analyzer samples from different bearings and record the operational status. The Bearing Checker has a 1.5 mm headphone jack as shown in. The computer's sound card has a corresponding audio input. So the wire with nail jack 1.5 mm was connecting the output of the Bearing checker to the input of the computer sound card. In this the transfer of sound from the Bearing checker to the computer. The registration and storage of sound was performed using the free program Audacity. The sound was saved in mp3 format with Bit rate 128kbps. To control the audio recording earphones were used which were connected to a computer.

Data acquisition of fuel-injector sounds was performed on fuel-injectors of a New Holland TN65N multipurpose tractor, three fuel-injectors controlled electronically, one healthy (fuel-injector1), one slightly damaged (fuel-injector2) and one audibly deviating from a healthy state (fuel-injector3).

Additionally, data acquisition of fuel-injector sounds was performed on fuel-injectors of a Zetor 7711 tractor, used for viticulture, four fuel-injectors controlled mechanically, fuel-injectors4-5-6-7 all deviating from healthy state. All

malfunctioning fuel-injectors needed cleaning to restore their functionality. Additionally, a newly installed fuel-injector8 was added for testing.

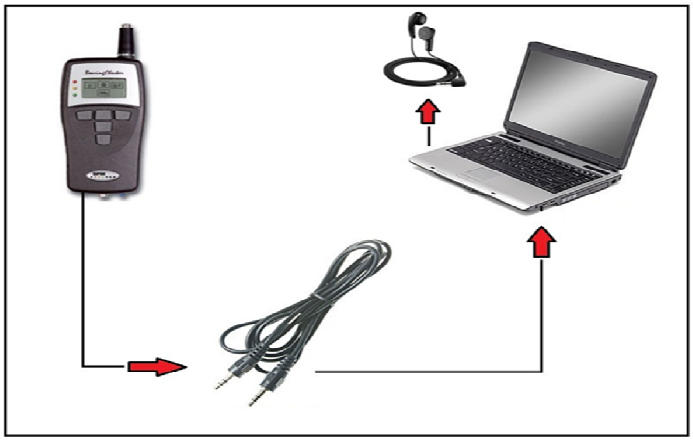

Fig. 3. Data acquisition setup for sounds emitted from malfunctioning fuel-injectors. The Bearing checker (by SPM Instrument) is shown on the left.

2.3 Signal processing and feature determination acquisition

To inspect raw vibration or sound signals, a wide variety of techniques have been introduced that may be categorized into two main groups: classic signal processing and intelligent systems. To make mention of a few, FFT, Wigner–Ville distribution, wavelets, blind source separation, statistical signal analysis, and their combinations are classic signal processing methods. Neural network based, genetic algorithm based, fuzzy logic based, various similar classifiers, expert systems, and hybrid algorithms can be classified as intelligent systems. Feature extraction was performed using seven features. The first six features were introduced in (Lei et al., 2009): Kurtosis, Skewness, Crest, Clearance, Shape and Impulse Indicators. A newly proposed feature consisting of the line integral of the acceleration or the sound signal is introduced in this work. All the used features provide statistical information about the nature of data, and were found to be reasonably good features for bearing fault detection. The Kurtosis is the fourth moment about the mean normalized with variance and for N points is given by Eq. 1. All other features are given by Eqs. 2-6.

$$Kurtosis = \frac{\sum_{i=1}^{N}(x_i - \mu_X)^4}{N\sigma_X^4} \quad (1)$$

$$Skewness = \frac{\sum_{i=1}^{N}(x_i - \mu_X)^3}{N\sigma_X^3} \quad (2)$$

$$Crest\ Indicator = \frac{\max|x_i|}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)^2}} \quad (3)$$

$$Clearance\ Indicator = \frac{\max|x_i|}{\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|x_i|}\right)^2} \quad (4)$$

$$Shape\ Indicator = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)^2}}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} \quad (5)$$

$$impulse\ Indicator = \frac{\max|x_i|}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}|x_i|}} \quad (6)$$

In Eqs. 1-6 μ_X and σ_X refer to mean value and standard deviation. The new line integral feature for *N* sampling points is given by Eq. 7:

$$\begin{aligned} LI &= \int_a^b ds \approx \sum_{i=1}^{N}\left\|\vec{r}(t_i+T_s)-\vec{r}(t_i)\right\| \\ &= \sum_{i=1}^{N}\sqrt{\left(x(t_i+T_s)-x(t_i)\right)^2+T_s^2} \\ &\approx \sum_{i=1}^{N}\left|x(t_i+T_s)-x(t_i)\right| \end{aligned} \quad (7)$$

Where N is the number of sample points (equal to 500) in the window used to calculate Kurtosis and the other features and the newly proposed line integral feature and Ts is the sampling period. The presented features were used for both the case of vibration signals from the gearbox test rig and the sounds collected from the fuel-injectors.

2.4 Self Organizing Map

The Self-Organizing Map also called SOM (Kohonen, 2001) is a neural network that maps signals from a high-dimensional space to a one- or two-dimensional discrete lattice (*M*) of neuron units. Each neuron stores a weight. The map preserves topological relationships between inputs in a way that neighbouring inputs in the input space are mapped to neighbouring neurons in the map space. SOM mimics the clustering behavior observed in biological neural networks, by grouping units that respond to similar stimuli together. Nerve cells, neurons, in the cortex of the brain seem to be clustered by their function. For example brain cells responsible for vision, form the visual cortex and those responsible for hearing form the auditory cortex.

The learning rule of the SOM consists of two distinct phases: when an input $\mathbf{x}$ is presented, search for the best matching unit or *bmu* through competition, and the update of the codebook patterns of the *bmu* and its neighbours. In the basic SOM the activations of the units are inversely proportional to their Euclidean distances from the input pattern hence the *bmu* can therefore be defined as:

$$b(\mathbf{x}) = \arg\min_{i \in M} \|\mathbf{x} - \mathbf{m}_i\| \quad (8)$$

where $b(\mathbf{x})$ is the index of the *bmu*, $\mathbf{m}_i$ is the codebook vector of unit i and $\mathbf{x}$ is the input pattern vector. The update part of the rule moves the *bmu* and its neighbours toward $\mathbf{x}$ to slightly enforce maps response to the pattern. The update rule can be written as follows:

$$\Delta\mathbf{m}_i = \gamma \cdot h(b(\mathbf{x}), i)(\mathbf{x} - \mathbf{m}_i) \quad (9)$$

where γ is a learning rate parameter and $h(b(\mathbf{x}), i)$ captures the neighborhood interaction between the *bmu* $b(\mathbf{x})$ and the unit i being updated. We can also write equation (9) as:

$$\Delta\mathbf{m}_i = \gamma \cdot H(\mathbf{x}, i)(\mathbf{x} - \mathbf{m}_i) \quad (10)$$

where $H(\mathbf{x}, i)$ is a shorthand notation for $h(b(\mathbf{x}), i)$. Equations (8) and (10) define a Hebbian learning rule, where the strength of the training step is determined not only by the learning rate parameter $0 < \gamma \leq 1$, but also by the relationship of the updated unit i with the *bmu* $b(\mathbf{x})$ on the map.

The inter-unit relationships are captured by the neighborhood $h(i, j)$ which defines how strongly units are attracted to each other. In essence the learning rule of the SOM defines the model as a collection of competitive units that are related through the neighborhood function. In practice, the units are placed on a regular low dimensional grid and the neighborhood is defined as a monotonically decreasing function on the distance of the units on the map lattice, thus creating a latent space, which has the dimension of the map grid and flexibility determined by the neighborhood function. The SOM can produce a flawless, in the sense that the map follows the manifold, embedding when the dimension of the map grid matches the dimension of the input data manifold. A typical choice for the neighborhood function is a Gaussian:

$$h(i, j) = \exp\left[\frac{d_M(i, j)^2}{2\sigma^2}\right] \quad (11)$$

where $d_M(i, j)$ is a distance measure in the map space (M), σ^2 is the variance of the Gaussian. The radius of the neighborhood is usually but not necessarily decreased during training. Likewise, the learning rate parameter γ is normally decreased in accordance to a predetermined cooling schedule, aiming to allow the map sufficient time and freedom to organize before fine tuning the codebook.

2.5 One Class SOM

In most cases of fault development in machinery there is no unique description of the faults but there are available a number of components that are either new or in different stages of malfunctioning behavior, which can not be quantified exactly. This is a common situation since the possible faults are either too many to reproduce or it is too costly to reproduce them. In some cases it is even impossible if the components that might experience a fault are involved in safety critical infrastructure. In safety critical applications, it is important to detect the occurrence of abnormal events as quickly as possible before significant performance degradation results.

Therefore, contrary to the approach followed for the cases where there are specific faults clearly defined, in usual cases only the healthy components can be used as target classification class and subsequently one-class classification methods are preferred. One-class classification has the following characteristics:

- Only information of target class (not outlier class) is available;
- Boundary between the two classes has to be estimated from data of only genuine class;
- Task: to define a boundary around the target class (to accept as much of the target objects as possible, to minimize the chance of accepting outlier objects).

As shown if Fig. 4, given a target domain X_T there are two errors that can be defined E_I related to false rejected target objects and E_{II} related to false accepted outlier objects. The circular area corresponds to the rough description of the target domain by the selected one class classifier.

Using a uniform outlier distribution also means that when E_{II} is minimized, the data description with minimal volume is obtained. So instead of minimizing both E_I and E_{II}, a combination of E_I and the volume of the description can be minimized to obtain a good data description.

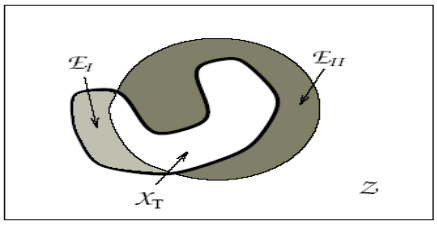

Fig. 4. Domains of target dataset and one-class classifier.

At first, an one class SOM (OCSOM) is trained with normal operation data. Then the feature vector corresponding to the unidentified measurement is compared with the weight vectors of all map units, and if the smallest difference exceeds a predetermined threshold, the process is probably in a fault situation. This conclusion is based on the assumption that a large quantization error corresponding to the operation point belonging to the space not covered by the training data. Therefore, the situation is new and something is possibly going abnormal. Depending on how far away the current process is deviating from the normal operation state, a quantitative degradation index can be calculated.

In the condition monitoring application, the one-class SOM (OCSOM) builds a model from training on healthy bearing and fuel-injector data and then classifies test data as either normal or outlier based on its geometrical deviation from the healthy training data. During novelty recognition, the unseen exemplar from a bearing or fuel-injector of unknown health state forms the input to the network and the SOM algorithm determines the best matching unit. In Saunders & Gero (Saunders and Gero, 2001) and Vesanto (Vesanto et al., 1998), if the vector distance or quantisation error between the best matching unit (bmu) and new exemplar data ($\mathbf{x}^{\mathbf{NEW}}$) exceeds some pre-specified threshold (d) then the exemplar is classified as novel. Eq. 12 gives the minimum vector distance for the bmu and compares this to the threshold.

$$\min(\sum_{j=0}^{n-1}(\mathbf{x}_j^{NEW} - \mathbf{m}_i)^2) > d,\ i \in M \qquad (\mathbf{12})$$

Where *M* represents the SOM grid of neurons as in equation (8).

There are many different heuristics to define a threshold depending on the utility of the threshold and the particular structure of the data set. A simple way to determine a threshold (d) relies on the distances between codebook vectors and target vectors in the training set that have selected them as bmu which is a measure of the quantization error. These distances have to be calculated first according to Eq. 13:

$$\text{distances} = \min(\sum_{k=0}^{N-1}(\mathrm{x}_k^{TARGET} - \mathrm{m}_i)^2),\ i \in M \qquad (\mathbf{13})$$

The threshold is determined according to the Matlab code given here which is further explained below:

```
distances_sorted=sort(distances);
frac=round(fraction_targets*length(target_set));
threshold=(distances_sorted(frac)+distances_sorted(frac+1))/2;
```

By selecting the threshold to represent a fraction of the distances for the whole training set we can get distance values representing the most proximal to the codebooks data vectors when the distances are sorted. In this case the quantisation errors might be due to outliers so the fraction error would represent the distances that were calculated for a distribution of the distances including outlier values. By taking

the 99% fraction of the distances between data and codebooks as belonging to the dataset we define a description hypersphere that has a radius including the 99% of the data. This leaves a 1% outliers that will be classified as such since they exceed the target set description radius. Corresponding to Fig. 4 this would be the contributing factor to E_I while we have minimized the target data description by thresholding according to a fraction of the data. In plain terms it means that by tightening the target data description we can afford to a number of false rejects in order to obtain a more accurate novelty detection which would be impossible with a very wide region of acceptance due to a very high threshold. In an explanatory schematic (Fig. 5) one can see the different areas defined by the threshold to the best matching units and the Voronoi polygons defining the domains of the OCSOM neurons. It can be seen that some data points that would be classified as belonging to a neuron now fall outside the threshold-defined polygon that delimits the target data from the novel data belonging to vibration and acoustic signatures from damaged components (this is just an illustration, the actual data are high dimensional and can not be visualized directly).

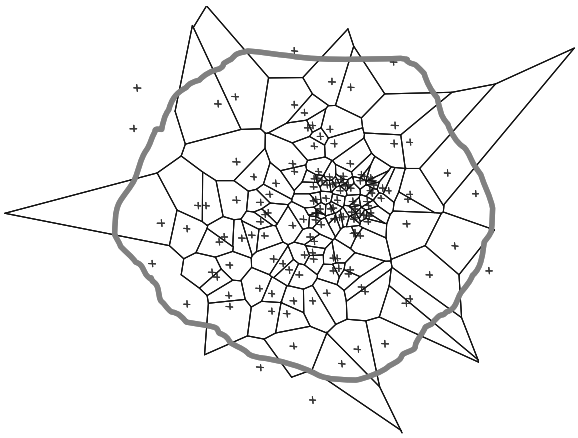

Fig. 5. Domains of target dataset and associated Voronoi polygons and threshold based classifier for OCSOM. The threshold defined target data fall inside the grey border line.

3 Results and discussion

For the bearing fault recognition an OCSOM was used. A validation set was used to test the generalisation performance of the OCSOM. To test the effectiveness of OCSOM, the 75% of the target set containing only healthy bearing instances have been used for training while the 25% have been used in order to test the generalisation of the OCSOM. These were results for an OCSOM of 64 neurons (arranged in a 8x8 grid) which gave the best results by testing different sizes between 5 and 25. The implementation used the simulation software Matlab 2010b (Mathworks). Seven features of the same type from each accelerometer were used. The same order has been used for the horizontal accelerometer in order to build the

fusion vector. The fusion (by direct concatenation) of 14 vibration features from both the vertical and the horizontal accelerometer, due to their complementary nature, results in more accurate separation of classes regarding fault position as one can deduce from the results presented in Table 1. The correct novelty detection percentage for fusion reaches 94.31% which is higher than the results for both horizontal (67.65%) and vertical (90.02%) which means one accelerometer alone cannot detect fault presence accurately. The complementarity of features was expected because the vibration modes were measured in two orthogonal directions (vertical and horizontal) which carry projections of the vibration shapes on these independent axes. When using 25 neurons (5x5 grid) the false rejects decrease and a 97.35% correct healthy bearing recognition is achieved. At the same time the correct novelty detection percentage falls to 90.39%. So, overall, the 64-neuron architecture is better for novelty detection. This could be due to the complexity of the damaged bearing data due to incorporating two different damage types (inner and outer race fault). The added value of the newly introduced feature of line integral is proven from training and testing with and without the line integral features. In the case the line integrals of the vertical and horizontal accelerometer signals are omitted (12 out of 14 features kept), the result is 92.94% for healthy and 81.37% for novelty detection which is much less than when they are included (see Table 1, the result for fusion). When omitting only the line integral of the vertical accelerometer signal (13 out of 14 fusion features kept) the result is 94.51% for healthy and 84.31% for novelty detection. When omitting only the line integral of the horizontal accelerometer signal the result is 94.90% for healthy and 90.20% for novelty detection. So, the inclusion of the line integral feature enhances the results.

Table 1. Results of OCSOM with 64 neurons predicting bearing health condition.

Actual health state	Healthy bearing according to OCSOM	Extended fault according to OCSOM
Healthy bearing	96.08% (fusion) 96.86% (vertical) 94.90% (horizontal)	3.92% (fusion)
Extended fault	5.69% (fusion)	94.31% (fusion) 90.02% (vertical) 67.65% (horizontal)

The OCSOM was used to classify the fuel-injectors to a target class corresponding to healthy fuel-injectors and detect outliers indicating fuel-injectors that are malfunctioning. As target class, features belonging to fuel-injector1 have been used. All other fuel-injectors have been used for testing the performance of the OCSOM. The OCSOM was calibrated by splitting the data to 75% training of the target set containing only healthy bearing instances and 25% testing sets has resulted in 100% correct classification for the target class of fuel-injector1 and 99.65% (97.89% without using the feature of the line integral) when using fuel-injector7 as outlier class for testing. These were results for a OCSOM of 100 neurons (arranged in a 10x10 grid) which gave the best results by testing different sizes between 5 and 25.

Further testing of the obtained OCSOM classifier was performed using all available fuel-injectors. Results are shown in Table 2. It is evident that all fuel-injectors have been identified correctly based on their respective condition. The slightly damaged second fuel-injector has also been identified as midway to damage which is accurate according to the expert opinion based on the sound emission from that fuel-injector.

Table 2. Results of OCSOM based classification of fuel-injector health condition.

Fuel-injector no. #	Actual condition	OCSOM classifies as healthy (percentage)	OCSOM classifies as outlier (damaged)
1	Healthy	99.21	0.79
2	Slight damage	27.02	72.98
3	Damaged	1.75	98.25
4	Damaged	6.49	93.51
5	Damaged	9.65	90.35
6	Damaged	2.81	97.19
7	Damaged	1.32	98.68
8	Healthy	95.44	4.56

In safety critical applications of novelty fault detection it is important to establish what degree of change is significant. Normal system behaviour may shift, for example, due to aging, system modifications, seasonal changes and change in operating conditions. An important issue concerns obtaining robust novelty thresholds that lead to reliable novelty detection. Novelty detection algorithms based on one-class neural networks have to be trained with data which adequately span the operating envelope so that false positives would not occur during normal operation.

4 Conclusions

It has been shown that the OCSOM can perform data fusion from accelerometer sensors through combining vibration features. These features can be used to detect faults in roller bearings and can therefore prove to be a powerful tool for bearing health monitoring. Different bearing faults can be detected against healthy bearings with high accuracy by using the collective response of several features and the fusion of different sensors, which may not be obvious by just looking at the data using other diagnostic techniques. The use of several features and a newly introduced feature, the line integral of the acceleration signal has given promising results in detecting the position of bearing faults. The feature based fusion of the vertical and horizontal accelerometer signals has increased the accuracy of bearing fault detection to more than 95% (more than 96% for healthy and 94% for faulty bearings). In the case of fuel-injector malfunctioning detection, the same type of features has been used. Due

to the nature of the problem, relying only on acoustic signatures from healthy fuel-injectors, one-class classification has been used. A one-class SOM has been used and has given very promising results. Further it was possible to detect correctly the condition of all the fuel-injectors that were presented to the one-class SOM. This result indicates that OCSOM is a robust classifier and can be used for detecting fuel-injector malfunction with high confidence. It is planned that this work be extended to include more real data, different features and fault types for bearings and gear boxes and also different types of fuel-injectors. A further improvement of OCSOM could result from defining context sensitive thresholds and also activation profiles that could be implemented as a kernel map indicating novelty through neuron activity bursts. The presented OCSOM technique for novelty detection can be extended to other fields where activity monitoring and novelty detection are needed, like process control, network security and sensor networks for various applications.

Acknowledgments. The authors would like to thank Greek Scholarships Foundation.

References

1. Kohonen, T. (2001) Self Organizing Maps, 3rd ed. Heidelberg: Springer.
2. Lei, Y., He, Z., Zi, Y. (2009) Application of an intelligent classification method to mechanical fault diagnosis. Expert Systems and Applications, 36, p.9941-48.
3. McFadden, P. D. and Smith, J. D. (1984) Vibration monitoring of rolling element bearings by the high-frequency resonance technique – A review. Tribology International, 17(1), p.3–10.
4. Paya, B. A., Esat, I. I. and Badi, M. N. M. (1997) Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. Mechanical Systems and Signal Processing, 11(5), p.751–65.
5. Saunders, R. and Gero, J. S. (2001) A Curious Design Agent: A Computational Model of Novelty-Seeking Behaviour in Design. In Proceedings of the Sixth Conference on Computer Aided Architectural Design Research in Asia (CAADRIA 2001). Sydney, p.345-50.
6. Sawalhi, N. (2007) Diagnostics, Prognostics and Fault Simulation for Rolling Element Bearings, PhD Thesis, University of New South Wales, Australia.
7. Vesanto, J., Himberg, J., Siponen, M. and Simula, O. (1998) Enhancing SOM Based Data Visualization. In Proceedings of the 5th International Conference on Soft Computing and Information/Intelligent Systems. Methodologies for the Conception, Design and Application of Soft Computing, Vol. 1, p.64–67, Singapore: World Scientific.
8. Xu, Z., Xuan, J., Shi, T., Wu, B. and Hu, Y. (2009) Application of a modified fuzzy ARTMAP with feature-weight learning for the fault diagnosis of bearing. Expert Systems with Applications, 36, p.9961-68.