Condition monitoring of Mechanical Subsystems of Agricultural Vehicles Based on Fusion of Vibration Features

D. Moshou[1], D. Kateris[1], N. Sawalhi[4], I. Gravalos[2], S. Loutridis[3], T. Gialamas[2], P. Xyradakis[2], Z. Tsiropoulos[2]*

[1]Aristotle University, School of Agriculture, Agricultural Engineering Laboratory, Thessaloniki 54124, Greece, E-mail: dmoshou@auth.gr
[2]Technological Educational Institute of Larissa, School of Agricultural Technology, Department of Biosystems Engineering, Larissa 41110, Greece
[3]Technological Educational Institute of Larissa, School of Technological Applications, Department of Electrical Engineering, Larissa 41110, Greece
[4]School of Mechanical and Manufacturing Engineering, University of New South Wales, Sydney 2052, Australia

e-mail of corresponding author: dmoshou@auth.gr

Summary

Bearings form an essential part of the mechanical subsystems of agricultural machinery and their failure is one of the most common causes of machine breakdowns. Accordingly, in order to increase reliability and reduce loss of production, condition monitoring of bearings has become more and more important in recent years. The use of vibration signals is quite common in the field of condition monitoring and fault diagnosis of bearings. Vibration analysis is based on the use of acceleration measurements from bearings in order to assess their health status. Advanced signal processing is used to construct a number of informative features that are sensitive to specific bearing faults and fault locations. The fusion of specific statistical features and the introduction of new features enable the accurate discrimination of faults based on their location. The capability of neural networks to visualize high-dimensional data is well known. They map nonlinear statistical relationships among variables of high dimensional input data on a low dimensional network, in a topology preserving fashion. The presented work concerns a neural network based diagnostic system architecture for monitoring the topology of extended faults in bearings. A Multilayer Perceptron (MLP) based approach has been used to map time series of feature data produced by acceleration sensors in order to capture the process dynamics. The evolution of system states following the bearing health trend has been shown to successfully identify different bearing faults according to their location.

Key word: bearings, condition monitoring, vibrations, neural networks, sensor fusion

Introduction

Bearings are located at the heart of rotating machinery and play a very important role in industrial applications and are mainly used to support and fix the axle in rotating machinery. Their failure in practical operation can lead to the breakdown of the whole machine. Accordingly, to increase reliability and reduce loss of production condition monitoring of bearing gets more and more important in recent years. 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).

Material and methods

Today's industry uses increasingly complex rotating machines, some with extremely demanding performance criteria. Bearings are located at the heart of rotating machinery and play a very important role in industrial applications and are mainly used to support and fix the axle in rotating machinery. Their failure in practical operation can lead to the breakdown of the whole machine. Accordingly, to increase reliability and reduce loss of production condition monitoring of bearing gets more and more important in recent years.

Rolling element bearings often fail due to spall and extended defects in the inner and outer races and the rolling elements. Attempting to diagnose faults in these systems is often a difficult and daunting task for operators and plant maintainers. Machine failure can lead to economic loss and safety problems due to unexpected and sudden production stoppages. In rotating machinery, the root cause of faults is often faulty rolling element bearings. One way to increase operational reliability and thereby increase machine availability is to monitor incipient faults in these bearings.
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 the 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 Figure 1).

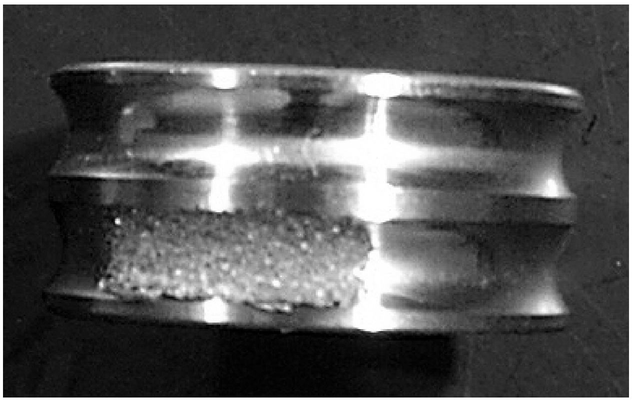

Figure 1. Example of an extended fault in the inner race.

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 Figure 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.

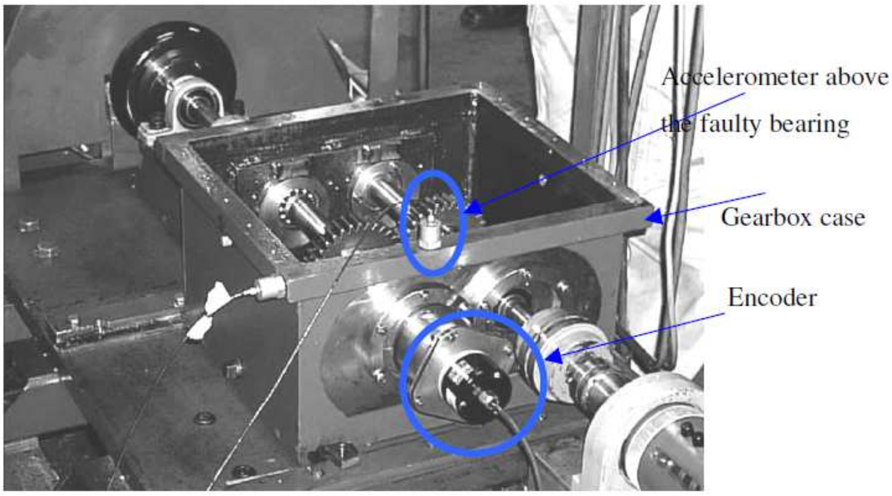

Figure 2. The spur gear rig.

Signal processing and feature determination

A machine vibration signal is composed of three parts, stationary vibration, random vibration, and noise. To inspect raw vibration 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. Currently, industrial applications of intelligent monitoring systems have increased due to the progress of 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 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 a sliding window of N sampling points is given by Equation (1). All other features are given from Equations (2)-(6).

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

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

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

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

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

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

In equations (1)-(6) the symbols μ_X and σ_X refer to mean value and standard deviation.
The new line integral feature for a sliding window of N sampling points is given by Eq (7):

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

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. Given the high sampling rate of 48 kHz and the domination of the signal from high frequencies (especially due to the presence of faults), the final approximation contains only acceleration values. The feature vectors are then fed to the MLP for training.

Neural network approach for monitoring fault topography

Feed-forward neural networks (Rumelhart et al., 1986) provide a general framework for representing non-linear functional mappings between a set of input and output variables. The multilayer perceptron is a type of neural network that posseses the universal approximation property (Bishop, 2005). A general structure is shown in Figure 3. The input to this network is the feature vector extracted from the object to be classified, and the output is typically a block code where one output is high indicating the class of the object and all other outputs are low. The weights connecting the nodes are determined using some training rule with a set of feature vectors, the training set.

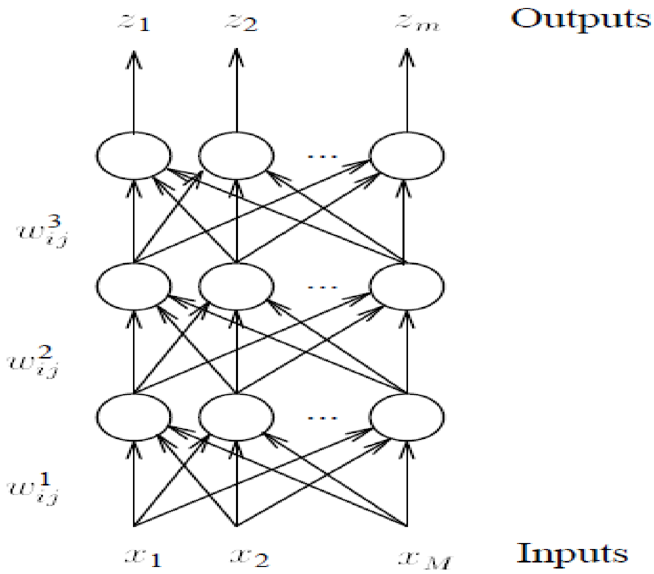

Figure 3. Multilayer perceptron. Superscripts are used to denote layers. Layer 1 is the first hidden layer and the inputs can be considered as Layer 0.

Three layer networks (one hidden layer) are universal approximators. However, this does not directly impact the classification problem. In fact an MLP with two hidden layers carries the property of forming arbitrary decision surfaces which makes it a better candidate for complex classification problems. In the case of open decision regions, an MLP with one hidden layer as in the current case of feature-based monitoring will suffice. The decision regions with respect to network architecture are shown in Figure 4.

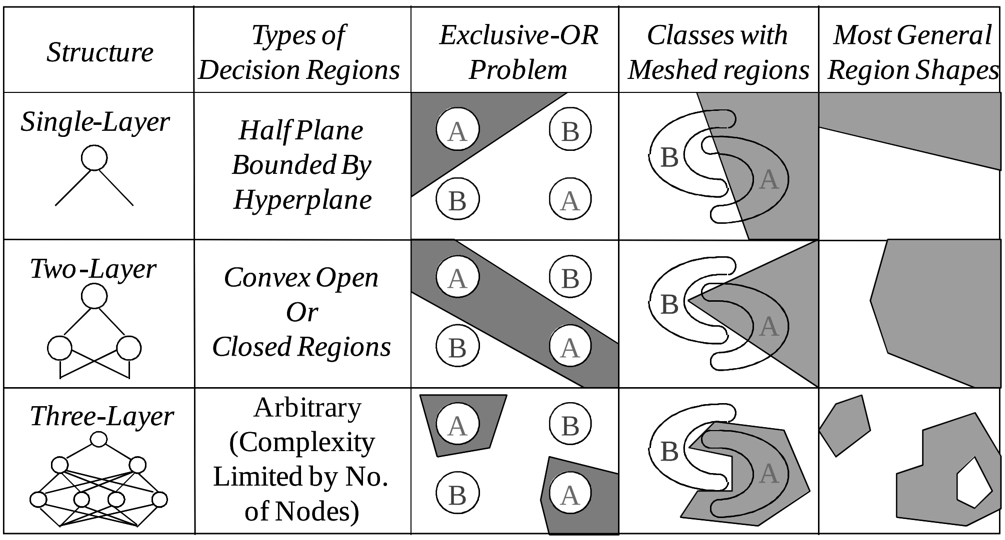

Figure 4. Decision regions of MLP according to the architecture and number of hidden layers.

Bayesian regularisation (Bishop, 1995) has been used to train the MLP in order to build a more robust classifier. It maps combinations of different characteristic features into fault categories, without performing an explicit modelling of the system. The MLP trained with Bayesian regularisation is a nonlinear projection method. The feature selection stage is one of the most important factors in the success of modeling. Once the MLP is trained, it is exposed to actual data from the system (in the form of the same type of features used for the training set) representing a yet unknown state. The data points are mapped onto the network as they are sequentially fed to the map describing the current state of the system. The current state MLP based estimation can be retrieved based on already stored labels representing fault classes.

A neural network that is designed to generalize well will produce a correct input-output mapping even when the input is slightly different from the examples used to train the network. When, however, a neural network learns too many input-output examples, the network may end up memorizing the training data. This phenomenon can be produced in case the network discovers a feature (due to noise, for example) that is present in the training data but it is not representative feature of the underlying function that has to be modelled. Such a phenomenon is referred to as “overfitting” or “overtraining”. When the network is overtrained, it loses the ability to generalize between similar input-output patterns. There exist two most common methods to avoid overfitting: early stopping and weight decay (Plaut et al. 1986). Early stopping has the advantage of being quick, since it shortens the training time, but it is not making full use of the available data apart from the fact that it is poorly defined. On the other hand, weight decay is well defined but quite time consuming.

In order to make the learning task well posed (or at least less ill-posed) a common solution is to introduce regularizers. That is, instead of only minimizing an error of fit measure like the commonly used mean square error shown in (8) the error functional takes the form shown in (9). The additional regularization term $E_w = \lambda R(\mathbf{W})$ expresses prior beliefs about the solution.

$$E(\mathbf{W}) = \frac{1}{2N}\sum_{n=1}^{N}[y(n) - f(\mathbf{W}, x(n))]^2 \qquad (8)$$

$$E(\mathbf{W}) = \frac{1}{2N}\sum_{n=1}^{N}[y(n) - f(\mathbf{W}, x(n))]^2 + \lambda R(\mathbf{W}) = E_D(\mathbf{W}) + E_w(\mathbf{W}) \qquad (9)$$

Where λ is the regularization parameter which weighs the importance of $R(\mathbf{W})$ relative to the error of fit $E_D(\mathbf{W})$. The effect of the regularization term is to shrink the model family F, or make some models more likely than others. As a consequence, solutions become more stable to small perturbations in the training data.

Another way to parametrize (9) is the following:

$$E(\mathbf{W}) = F(\mathbf{W}) = \beta E_D(\mathbf{W}) + \alpha E_W(\mathbf{W}) \qquad (10)$$

The optimal values of α and β can be calculated at the minimum point of a Bayesian formulation of the posterior probability of network weights with respect to data (Bishop, 2005):

$$\alpha^{MP} = \frac{\gamma}{2E_W(\mathbf{w}^{MP})} \text{ and } \beta^{MP} = \frac{n-\gamma}{2E_D(\mathbf{w}^{MP})} \qquad (11)$$

Where $\gamma = N - 2a^{MP} tr(\mathbf{H}^{MP})^{-1}$ is called the effective number of parameters ($\mathbf{H} = \beta\ \nabla^2 E_D + \alpha\ \nabla^2 E_W$ is the Hessian matrix of the objective function), and *N* is the total number of parameters in the network while *n* is the number of training data. The usefulness of parameter γ is that it provides a measure of how many parameters in the neural network are effectively used in reducing the error function. It can take values between zero and N.

Bayesian optimization of model control parameters has three important advantages:
1. No “test set” or “validation set” is involved, so all available training data can be devoted to both model fitting and model comparison.
2. Regularization constants can be optimized on-line, i.e., simultaneously with the optimization of ordinary model parameters.
3. The Bayesian objective function is not noisy, in contrast to a cross-validation measure.

An example of estimating a sinusoid function given noisy data is given in Figure 5. The non-regularized network overfits the data by memorizing the noise. The prediction of the network that has used Bayesian regularization is shown with a dotted line.

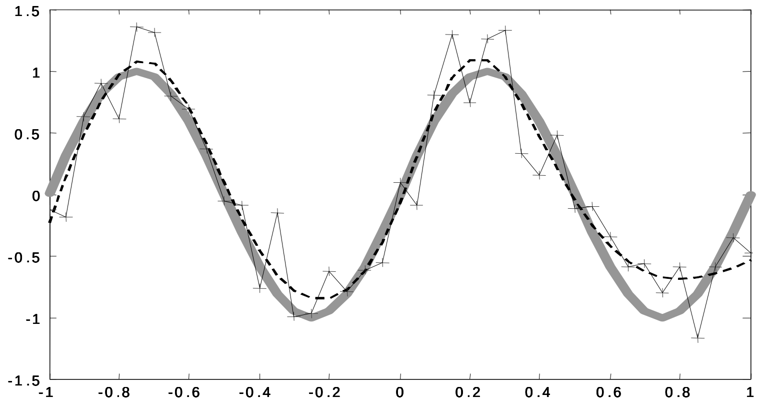

Figure 5. A BR trained network (dotted line). Training data are indicated by crosses while the smooth curve represents the function that has to be approximated.

Data fusion

The data fusion approach combines data from multiple sensors (and associated databases if appropriate) to achieve improved accuracies and more specific inferences that could not be achieved by the use of only a single sensor (Hall, 1990). This concept is hardly new:- living organisms have the capability to use multiple senses to learn about the environment. The brain then fuses all this available information to perform a decision task.

One of the first definitions of data fusion came from the North American Joint Directors of Laboratories (JDL) (White, 1990; Worden and Dulieu-Barton, 2004), who define data fusion as a:- multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation and combination of data from single and multiple sources. We can think of the heterogeneous sensors monitoring a certain process as being “windows” into the phenomenon under observation. Sensors can either have their own window, or the windows “overlap” in space or time. This way, the information obtained can be thought of as “decomposed” or “fragmented” by the sensors, which is sometimes called sensor fission (Dasarathy, 1997), and is related to so–called sufficient information (whether the character and number of sensors can indeed describe the phenomenon). The information fragments coming from sensors are exposed to spectral shaping, saturation, and noise; data fusion aims at retrieving the “interesting” characteristics of the phenomenon.

Development issues for data fusion systems depend on the phenomena that are observed, the type of sensors utilized, and the inferences sought. These inferences in turn determine the type of techniques required. Generally, applications aimed at higher-level inferences require techniques from the artificial intelligence domain such as expert systems, template matching, neural networks and fuzzy logic. In the current condition monitoring application seven features from each accelerometer are combined in a fusion vector by direct concatenation and then they are fed to the MLP which determines iteratively through Bayesian regularisation the relevance of each feature to the classification result and accordingly maps the fusion feature vector to decision regions corresponding to fault categories.

Results and Discussion

For the experiments a multilayer perceptron with one hidden layer having 20 neurons was used. A validation set was used to test the generalisation performance of the neural network. To test the effectiveness of MLP, the 75% have been used for training while the 25% have been used in order to test the generalization of the MLP. The implementation used the Neural Net Matlab Toolbox (Mathworks). The number of neurons in the input layer was equal to the number of used features. Different numbers of neurons in the hidden layer were used, varying between 5 and 25 at steps of 5. Best results were obtained using a one hidden layer neural network with 20 neurons in the hidden layer, 14 inputs and three outputs (healthy - inner race - outer race) corresponding to fault position. The hidden layers and the output layer of the MLP had sigmoid neurons. The training algorithm of Bayesian regularisation has the advantage of avoiding overtraining hence improving generalisation performance. Adding more neurons in the hidden layer does not improve the generalisation result when Bayesian regularisation is used as the training algorithm because weights that are not used remain small, therefore, the extra hidden neurons become inactive. Hence, the preferred neural network configuration was the one with the lowest complexity from the validated ones that did not compromise performance.

Seven features of the same type from each accelerometer as shown in Table 1 according to order of presentation to the MLP. The same order has been used for the horizontal accelerometer in order to build the fusion vector.

Table 1. Ordering of the features for the vertical accelerometer as presented to the MLP.

1	2	3	4	5	6	7
Kurtosis	Line integral	Crest	Clearance	Shape	Impact	Skewness

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 Tables 2-4 which indicate the superiority of the fusion based classification result. The complementarity of features was expected because the vibration modes are measured in two orthogonal directions (vertical and horizontal) which carry projections of the vibration shapes on these independent axes. The order of presentation of the features is the following:

Table 2. Results of classification of faults depending on their position by using a vertical accelerometer. The testing set used 25% of the data.

Vertical (% correct class estimate similar to fusion)		
100	0	0
0	96.8627	3.1373
0.3922	1.9608	97.6471

Table 3. Results of classification of faults depending on their position by using a horizontal accelerometer. The testing set used 25% of the data.

Horizontal (% correct class estimate similar to fusion)		
84.7059	11.7647	0.7843
14.1176	87.8431	0.7843
1.1765	0.3922	98.4314

Table 4. Results of classification of faults depending on their position by using a vertical and a horizontal accelerometer and the fusion features from both. The testing set used 25% of the data.

Fusion (% correct class estimate for healthy-inner race fault-outer race fault)		
100	0	0
0.392	99.608	0
0	0	100

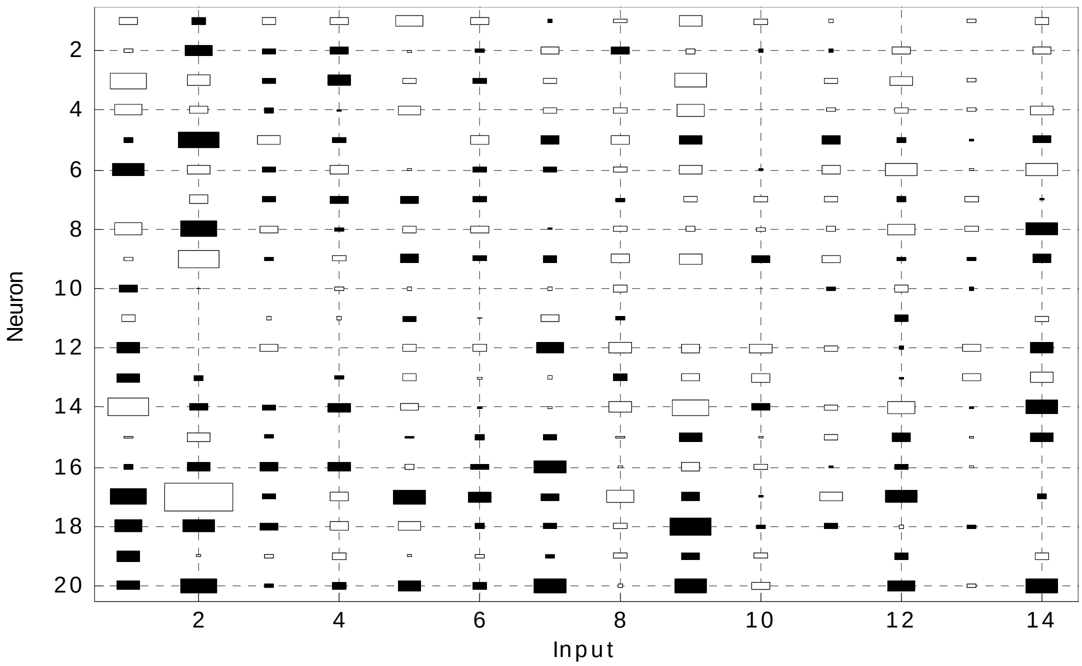

Figure 6. A Hinton diagram shows the values of the weights connecting vibration features and hidden neurons. White color indicates positive weight while black indicates a negative weight.

The importance of different features can be determined qualitatively through a Hinton plot of the weight matrix as shown in Figure 6. The contribution of the Kurtosis and Line integral of the vertical accelerometer (features 1 and 2) and the Line integral of the horizontal (feature 9) is significant compared to the other features. Also, the Shape Indicator and Skewness of both accelerometers (features 5, 7, and 12, 14) contribute to the excitation of the hidden neurons and subsequently to the resulting classification. Overall the Line integral shows the stronger contribution and from both accelerometers while the type of contribution is different per accelerometer indicating a complementary vibration component from each direction.

Conclusions

It has been shown that the MLP can perform data fusion from accelerometer sensors through combining vibration features. These features can be used to detect faults in roller bearings and discover the position of the faults, and can therefore prove to be a powerful tool for bearing health monitoring. Different bearing faults can be detected 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 kurtosis 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 acceleration signals has increased the accuracy of fault detection to 99% for different fault types. This result represented a substantial increase in discrimination performance of at least 10% for certain types of fault. It is planned that this work be extended to include more real data, different features and spall sizes for bearings in gearboxes or other machines.

References

Bishop, C., 2005. Neural Networks for Pattern Recognition, Oxford University Press, New York, USA, 504 pp.

Dasarathy, B.V., 1997. Sensor fusion potential exploitation – Innovative architectures and illustrative applications. Proceedings of the IEEE [85], 24–38.

Hall, D. L., 1992. Mathematical Techniques in Multisensor Data Fusion, Artech House, UK.

Lei, Y., He, Z., Zi, Y., 2009. Application of an intelligent classification method to mechanical fault diagnosis. Expert Systems and Applications [36], 9941-9948.

McFadden, P. D., Smith, J. D., 1984. Vibration monitoring of rolling element bearings by the high-frequency resonance technique – A review. Tribology International [17(1)], 3–10.

Paya, B. A.-Esat, I. I.-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)], 751–765.

Plaut, D., Nowlan, S., Hinton, G., 1986. Experiments on learning by backpropagation. Technical Report CMU-CS-86-126, Carnegie Mellon University, Pittsburgh, PA.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning internal representations by error propagation. In: Rumelhart D E, McClelland J L (Eds.), Parallel Distributed Processing (Part 1). MIT Press, Cambridge, USA, 318–362.

Sawalhi, N., 2007. Diagnostics, Prognostics and Fault Simulation for Rolling Element Bearings, PhD Thesis, University of New South Wales, Australia.

White, Jr., F.E., 1990. Joint directors of laboratories data fusion subpanel report. In: Proceedings of the Joint Service Data Fusion Symposium, DFS–90. 496–484.

Worden, K., Dulieu-Barton, J.M., 2004. An overview of intelligent fault detection in systems and structures. Structural Health Monitoring [3], 85–98.

Xu, Z.-Xuan, J.-Shi, T.-Wu, B.-Hu, Y., 2009. Application of a modified fuzzy ARTMAP with feature-weight learning for the fault diagnosis of bearing. Expert Systems with Applications [36], 9961-9968.