

# The Detection of Rolling-Element Bearing Faults in Non-stationary Quasi-Parallel Machinery Using Residual Analysis Augmented by Neural Networks

Dustin Helm<sup>1</sup>, and Markus Timusk<sup>2</sup>

<sup>1,2</sup>*Laurentian University, 935 Ramsey Lake Road, Sudbury P3E2C6, Canada*

*dm\_helm@laurentian.ca*

*mtimusk@laurentian.ca*

## ABSTRACT

This work proposes a methodology for the detection of rolling-element bearing faults in quasi-parallel machinery. In the context of this work, parallel machinery is considered to be any group of identical components of a mechanical system that are linked to operate on the same duty cycle. Quasi-parallel machinery can further be defined as two components not identical mechanically, but their operating conditions are correlated and they operate in the same environmental conditions. Furthermore, a new fault detection architecture is proposed wherein a feed-forward neural network (FFNN) is utilized to identify the relationship between signals. The proposed technique is based on the analysis of a calculated residual between feature vectors from two separate components. This technique is designed to reduce the effects of changes in the machines operating state on the condition monitoring system. When a fault detection system is monitoring multiple components in a larger system that are mechanically linked, signals and information that can be gleaned from the system can be used to reduce influences from factors that are not related to condition. The FFNN is used to identify the relationship between the feature vectors from two quasi-parallel components and eliminate the difference when no fault is present. The proposed method is tested on vibration data from two gearboxes that are connected in series. The gearboxes contain bearings operating at different speeds and gear mesh frequencies. In these conditions, a variety of rolling-element bearing faults are detected. The results indicate that improvement in fault detection accuracy can be achieved by using the additional information available from the quasi-parallel machine. The proposed method is directly compared to a typical AANN novelty detection scheme.

Dustin Helm et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.36001/IJPHM.2021.v12i2.2915>

## 1. INTRODUCTION

The early and reliable detection of incipient faults in machinery is crucial for the feasibility of a condition-based maintenance program. The detection of these faults is often accomplished using vibration-based methods. Signal processing algorithms work to improve the signal to noise ratio (SNR) of raw signals and extract a set of useful features that best describe the current state of the machine. The classification step uses these features to classify the machine as either faulted or healthy (or diagnose the type of fault present). One challenge in using vibration-based techniques is their application to machinery that has time-varying operational characteristics (nonstationary). These characteristics are most commonly the speed and load of the machine, however other factors such as ambient temperature, variables from human operators, process parameters and many others can also be considered. When a machine operates in a nonstationary manner, the measured vibration for the condition monitoring system undergoes frequency and amplitude modulation. These effects can be most clearly observed in the frequency variation of machine vibrations with respect to changing rotational speed or in cases where the signal is roughly modulated by the power delivered to the system. To illustrate this in an industrial context Figure 1 demonstrates the amplitude and frequency modulation of a gearbox of an electromechanical mining excavator. The lower half of the diagram represents the rotating speed of the gearbox.

This variability in the raw vibrations is not tied to the health or condition of the machine and can hinder the signal processing steps, it can also create large variations in the features that can cause the boundaries between classes to overlap and reduce the classification accuracy. Many techniques have been developed to deal with nonstationary signals to be applied to both the classification and the signal processing steps. For example, order tracking is a widely accepted technique for demodulating the changes in

frequency with respect to angular speed that is often integrated directly into hardware (Randall & Antoni, 2011). However, this technique requires accurate measurements of speed which can be difficult to implement, particularly in industrial applications. Another technique is the treatment of signals as so called cyclo-non-stationary signals to deal with the interaction of time and angle dependent factors in the signals (Abboud et al., 2016). Alternatively, signals can be analyzed in the time-frequency domain using techniques such as wavelet analysis or empirical mode decomposition (Lei, Lin, He, & Zuo, 2013). As effective as these techniques have been shown to be at increasing the SNR in cases where there are mild fluctuations in speed and load, there are a multitude of applications where machinery with widely fluctuating duty cycles create signals that cannot consistently be treated using these methods.

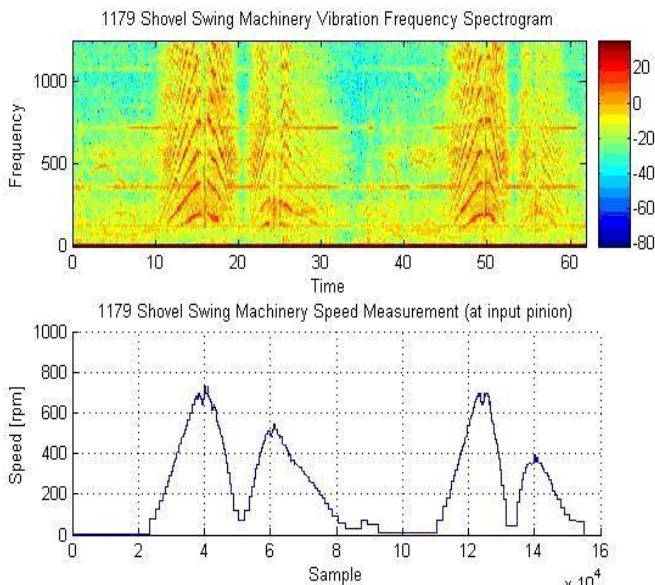


Figure 1. Vibration Response of Excavator Gearbox under Unsteady Operating Conditions

Another key technique for the detection of faults in machinery is using a method known as analytical redundancy relations (ARRs) (Staroswiecki & Comtet-Varga, 2001). Analytical redundancy is a case where there are two possible ways to determine a variable, and one of which is in the form of an analytical model (Isermann & Ballé, 1997). When using ARR as a basis to detect faults, the difference, or residual between the two estimates of a variable (measured and model-based) can serve as an indicator for the fault. In more complex systems with multiple possible faults, the ARR can be structured such that they contain fault diagnostic information (Gertler, 1997). In this case the number of possible ARR is equal to the number of sensors on the machine. Recently Gor et al. used ARR for fault accommodation in quadruped robots (Gor, Pathak, Samantaray, Yang, & Kwak, 2018), and

Willersrud et al. used ARR to detect faults during oil and gas drilling (Willersrud, Blanke, & Imsland, 2015).

Artificial Neural Networks (ANNs) provide an excellent solution for the classification of features from machine vibrations in condition monitoring systems. This is due to the ANN's ability to deal with the noisy and incomplete data sets that are typical of condition monitoring applications. It is often difficult to obtain complete representations of machine vibrations for every faulted condition across all operating states. ANN-based classifiers have been applied to bearing and gear fault detection in stationary machinery using statistical features with great success (Samanta, 2004; Samanta & Al-Balushi, 2003). More recently ANNs have been used along with time-frequency domain techniques to detect and diagnose faults (Barakat, Druaux, Lefebvre, Khalil, & Mustapha, 2011; Bin, Gao, Li, & Dhillon, 2012; Xie & Zhang, 2017). Strdczkiewicz and Barszcz demonstrated that by utilizing a backpropagation ANN and simple statistical features (RMS and peak-to-peak) it is possible to detect incipient faults in highly non-stationary wind turbine gear boxes (Strdczkiewicz & Barszcz, 2016). A good review of the application of machine learning and artificial intelligence to machine fault detection can be found in (Liu, Yang, Zio, & Chen, 2018).

Recent trends have seen an increase in the application of deep learning (DL) approaches to condition monitoring problems. Due to the rapidly growing ability of computational and data collection systems DL approaches are becoming more practical for industrial applications. DL approaches are a powerful tool for industry as they eliminate the need for application specific feature extraction techniques. Examples of DL applied to condition monitoring can be found in (Jia, Lei, Lin, Zhou, & Lu, 2016; Jiang, Wang, Shao, & Zhang, 2017; Zhao et al., 2019).

Auto-associative neural networks (AANNs), sometimes also referred to as autoencoders, are a specific type of neural network that are trained to reconstruct the input at the output (Kramer, 1992; Kramer, 1991). The key feature of the AANN structure is a bottle-neck in the center that forces the network to compress the data into a number of principal components that contain as much of the necessary information as possible for reconstruction. AANNs have been receiving much attention for their application to DL referred to as the deep auto-encoder (DAE). The typical implementation of a DAE involves sending raw sensor data into the input layer, passing it through a lower dimensional hidden layer and reconstructing the original data on the output layer (i.e. encoding then decoding the data). Principi et al. demonstrated that unsupervised deep autoencoders could outperform one-class support vector machines for detection of electric motor faults (Principi, Rossetti, Squartini, & Piazza, 2019). However, it has also been shown that DAEs have some difficulty representing the

noisy non-stationary signals common in fault detection (Haidong, Hongkai, Xingqiu, & Shuaipeng, 2018; Shao, Jiang, Zhao, & Wang, 2017).

AANN's can also be used a novelty detector for one class classification, wherein the difference between the input and output (referred to as the reconstruction error) indicates the likelihood of fault. Using AANNs to perform one class classification eliminates the need to train the system using data from the faulted condition. One of the first implementations of an AANN as a novelty detector for fault detection was done by Japkowitz et al. (Japkowicz, Myers, & Gluck, 1995), where the authors were able to detect faults in helicopter gearbox vibration signals. When the system remains healthy the reconstruction error is minimal because the input data closely matches the structure of the training data, however when a fault is present and the data changes, the networks reconstruction of the input will have significant error. This novelty detection approach allows the network to detect incipient faults without prior training on fault data that is often difficult to obtain. AANNs have been shown to be successful in detecting gear faults when coupled with wavelet analysis (Sanz, Perera, & Huerta, 2007). Using a priori information about fault signatures, multiple AANNs can also be configured and trained to classify fault types, in this framework AANNs have been shown to outperform other novelty detectors (Gianluca, Fromaigeat, & Etienne, 2016). AANNs have also been used for novelty detection for online tool wear monitoring, where the reconstruction error of the network output can indicate the presence and severity of tool wear (Wang & Cui, 2013). While the AANN when applied as a novelty detector removes the need for training with difficult to obtain fault data, it remains sensitive to the changes in the operational conditions of the machine. Changes in operating conditions will change the structure of the input data resulting variations in the output that could easily be interpreted as a fault. This results in a balance of sensitivity issue where thresholds must be set to balance between false positives and false negatives. This can be visualized in well-known ROC (receiver operator characteristic) curves.

Experimental feature residual analysis, first proposed in (Helm, Rose, & Timusk, 2016; Rose, Helm, & Timusk, 2016) and further investigated in (Helm & Timusk, 2017; Helm & Timusk, 2019) is a method for detecting faults in connected parallel machinery by analyzing the residual or difference between vibration features in the parallel subsystems. When a fault is present the residual between the features of the vibrations from the parallel subsystem will increase. This is revealed by thresholding the Euclidean distance between the feature vectors of each parallel subsystem. In the context of this method, connected parallel machinery is defined as identical mechanical subsystems that speed and load as well operating conditions at the same time (i.e., share the same forcing functions). By exploiting the relationship between the parallel subsystems, it was

demonstrated that this method can reduce the fault detection systems sensitivity to non-stationary operation and improve classification results. This technique is similar to analytical redundancy as proposed in (Willersrud et al., 2015), however the redundancy is not in the form of a mathematical model but rather comes from the redundant hardware configuration.

Experimental feature residual analysis as defined in (Helm & Timusk, 2017; Helm & Timusk, 2019) is limited in possible applications; this work looks to expand the possible applications to quasi-parallel machinery. Quasi-parallel machinery, unlike connected parallel machinery does not require the individual subsystems to be identical, nor do they have to have identical operating conditions, the only requirement is that they share a common forcing function. The operating conditions for each subsystem in quasi-parallel machinery will be related by some transfer function due to the shared forcing function.

The main contribution of this work is that it presents a new fault detection architecture that extends the work in (Helm & Timusk, 2017; Helm & Timusk, 2019) to be able to include connected machinery that does not necessarily operate in a perfectly parallel manner (quasi-parallel). This is accomplished through the addition of an FFNN to the experimental feature residual analysis technique to allow the utilization of the real time information from a connected but not strictly identical component to reduce the sensitivity of the system to fluctuations in speed and load. In this application the FFNN is setup to mimic the typical application of an AANN with the difference that the network is trained to reproduce the corresponding data from another subsystem rather than the input data. This allows the parallel method to be applied to a much wider range of industrial machinery.

This work will focus on applying experimental feature residual analysis to gearboxes connected in series. This arrangement for components can be considered quasi-parallel due to the relationship between the speed and load of the gear sets. Gearboxes are machine components that are critical for the transmission of power between actuators and loads. Gearboxes are used to change the speed and output torque of the machine. Consider the signal model for the vibrations of a healthy gearbox presented in (Abboud, Antoni, Sieg-Zieba, & Eltabach, 2017) (with an added term in the modulation function for variable load) which is shown in Eq. (1).

$$X(t) = d(t) + r(t) + b(t), \quad (1)$$

where  $d(t)$  is the deterministic component shown in Eq. (2),  $r(t)$  is the random component given in Eq. (3) and  $b(t)$  is the background noise given in Eq. (4).

$$d(t) = [M(\omega(t), L(t)) \sum_i a_i \cos(z_1 i \theta(t) + \varphi_i)] \otimes H_g^d \quad (2)$$

$$r(t) = (M(\omega(t), L(t))W(t) * \sum_i a_i \cos(\frac{z_1}{z_2} i \theta(t) + \varphi_i)) \otimes H_g^r \quad (3)$$

$$B(t) = W(t)M(\omega(t), L(t)) \otimes H_b, \quad (4)$$

where  $M$  is the modulation function,  $\omega(t)$  is the input speed,  $L(t)$  is the load (input torque),  $a_i$  and  $\varphi_i$  are the amplitude and phase of the  $i$ <sup>th</sup> Fourier coefficients respectively,  $z_1$  and  $z_2$  are the number of teeth on the input and output gears respectively,  $W(t)$  is white noise with unit standard deviation and each  $H$  is a linear time invariant (LTI) system that represents the signal transfer path for the different parts of the signal.

When two gearboxes are connected in series, they both produce vibrations based on this model, however there are some strict relations between the two given that they are connected. These are given in Eqs. (5-9).

$$\omega_1(t) = \omega_2(t) * Nn \quad (5)$$

$$L_1(t) = L_2(t)/N \quad (6)$$

where  $N = z_1/z_2$  (gear ratio of first gearbox), and since  $\theta(t) = \int_0^t \omega(t) dt$ , then  $\theta_1(t) = \theta_2(t) * N$ .

All the components of the signals that are not time invariant will be strictly related between two gearboxes. Therefore time-invariant relationships can be identified between the two signals by using a technique such as a neural network. It is this relationship that the neural network used for the system difference identification step attempts to model. However, when there is a change in the structure of either signal (i.e., when a fault is present), the identified relationship will no longer hold true.

## 2. METHODOLOGY

The technique that is developed for the detection of faults in quasi-parallel machinery is defined here. The key point is that the two quasi-parallel subsystems share some relation between their operational states due to being linked mechanically or electrically (i.e. changes in operation are related by some transfer function). This relation between the subsystems operating conditions leads to relations between their vibrations which are most visible in the feature domain. The feature domain here refers to the domain of values (features) that are extracted from the raw time domain signals that are likely to trend with a fault. Using a neural network, the system difference in the feature domain can be identified for healthy cases. By applying experimental feature residual analysis to the vibration signals and incorporating the neural network for one of the parallel data streams to adapt the data according to the identified system difference, faults can then be detected in either subsystem.

Figure 2 is a block diagram of the computational steps for this method in a general case. First signals from the

mechanical system to be monitored must be collected. Typically, transducers such as accelerometers are used however other signals such as temperature, noise, electric current or the control signals to the machine could also be used to indicate the health of the system.

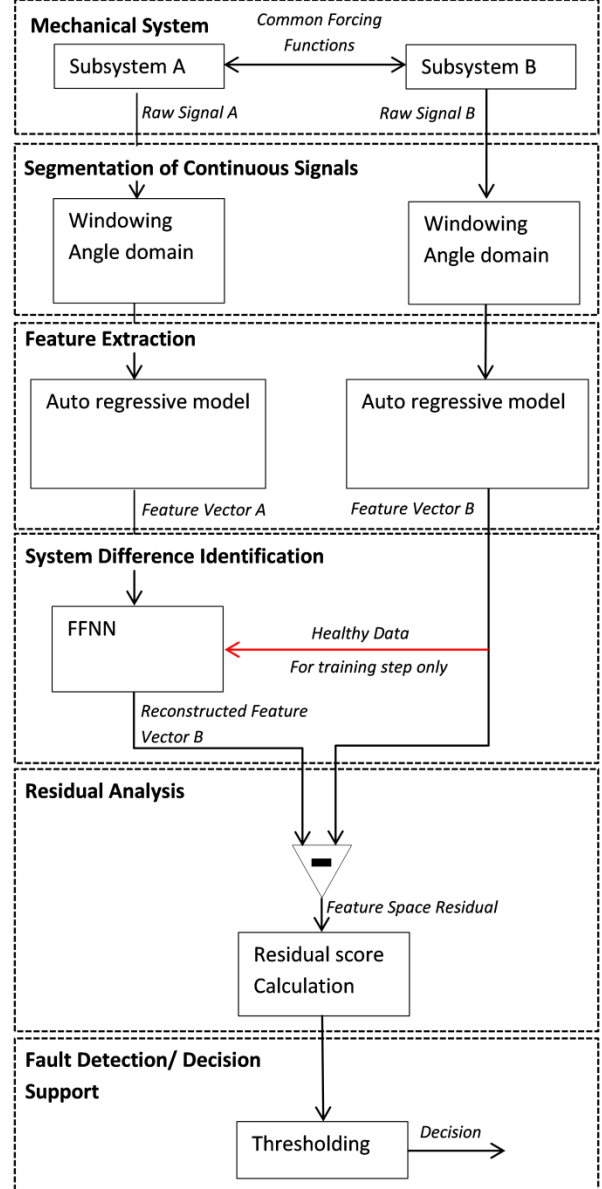


Figure 2. Signal Flow Diagram (red lines for training)

### Segmentation

The raw continuous time series signals are segmented into short samples to be analyzed. The segments can be in even time increments or even angle increments (corresponding to the rotation speed of the system). A windowing function can be applied to each segment depending on requirements for feature extraction. When a windowing function other than a rectangular window is used (ex. Hanning window) the segments should be overlapped to eliminate loss of data. In

this work, the signals were segmented at 8 shaft revolutions on the load side of the system. The segments were windowed with a Hanning window and overlapped by 50%.

### Feature Extraction

In this step several features are calculated from each segment to represent key indicators of faults in the signals. Any number of commonly employed features could be utilized, such as frequency domain envelope features, basic statistical features or time frequency domain features (e.g. wavelet domain). However, in this work well-known autoregressive (AR) model coefficients are used to characterize the raw vibration segments. The AR method is attractive due in part to its computational efficiency making it amenable to industrial applications. An AR model was generated for each data segment that closely models the vibrations characteristics. The form for this model is given in Eq. (10) where  $y$  is the model output,  $x$  is the input and  $a_1$ - $a_p$  and  $b_0$  are the model coefficients. The models can be calculated using a linear least squares method that will minimize the total error between the model output and the raw data.

$$y(n) = (a_1y(n-1) + a_2y(n-2) \dots a_p y(n-p) + b_0x(n)) \quad (10)$$

After a model is created for a segment, the coefficients ( $a_1$ - $a_p$ ) are used as the signal features. These features have been shown to be sensitive to the health state of the machine (Cong, Chen, & Dong, 2012; Timusk, Lipsett, & Mechefske, 2008). AR models have also been demonstrated to be a useful tool for detection of fretting in non-stationary bearings (McBain, Lakanen, & Timusk, 2013). The order for the model ( $p$  in the given equation) determines the number of elements in the feature vector. The model order can be optimized using Akaike information criterion (AIC) (Figueiredo, Figueiras, Park, Farrar, & Worden, 2011; X. Wang & Makis, 2009). Other methods such as the one proposed by Chen and Mechefske could also be used (Chen & Mechefske, 2001). When evaluating the AIC over the same data used to fit the model, the AIC will always decrease as the model order increases (See an example in Figure 3).

However, for classification higher model orders are less desirable due to the higher dimensional feature space. In this work AR models of order 10 were deemed acceptable and not optimized further.

### Fault detection

The feature vectors calculated from each segment are used as inputs to the FFNN. The FFNN is trained using a set of healthy data to minimize the distance between feature vectors of each subsystem. This is done by using the vectors from one subsystem as inputs to the FFNN and the features calculated from the other subsystem at the same time segment as the targets to train the network. In doing so the

network is trained to take a feature vector from one segment and reproduce the features from the same time interval on the other channel. Once the network is trained, data of unknown class can be passed through the network and in cases where the difference between subsystems remains the same, the residual between the network output and the other subsystem will remain unaffected. However, when the system difference is changed (i.e., in the presence of a fault) the residual will increase. A single residual score can be calculated by simply taking the squared sum of the feature space residuals (Euclidean distance of the feature vectors). Other options for the residual score could be other types of distance metrics such as Chebyshev or cosine distance, however this function should be the same as the cost function that was used to train the network. This allows for the use of a simple threshold for separation between the healthy and faulted classes. The proposed neural network functions in a similar way to an AANN when used for novelty detection however the key difference is that it is reconstructing a data set from a quasi-parallel signal rather than the same data. The difference between the reconstruction and the actual data can be used to indicate the presence of a fault.

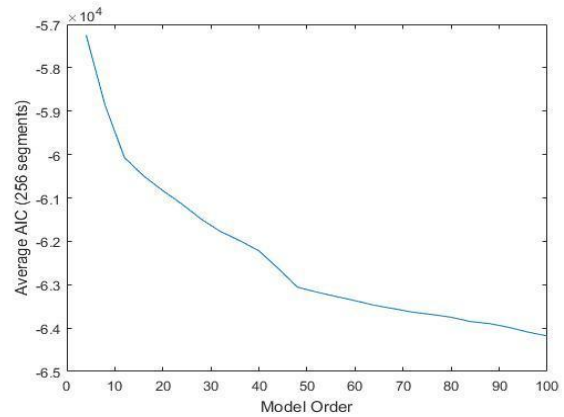


Figure 3. AIC with respect to model order

In this work, the FFNN used had 5 layers with 10, 12, 6, 12 and 10 neurons respectively. Several layouts were tested in the initial stages of this work and this layout was found to outperform the rest on the given test and validation data (see table 1), performance was evaluated based on the reconstruction error of the trained network. The chosen architecture generalized well without overfitting the data. While this architecture is not considered optimized the performance for this work was deemed acceptable and it was not optimized further. The transfer functions for the input and output layers were linear, while the rest were hyperbolic tangent sigmoid transfer functions, in order to model any possible non-linear relations. The network was trained using the Levenberg-Marquardt backpropagation algorithm as described by Hagan and Menhaj (Hagan & Menhaj, 1994). It was implemented using the Matlab Neural

Network toolbox. This training algorithm was chosen due to its performance on small networks (Hao & Wilamowski, 2011). The network was trained using data from a random set of four of the healthy tests. Data from a single test was used for validation to stop overfitting, while the rest were left for testing. The results using the parallel FFNN are compared to a standard AANN architecture. This network is setup the same as the FFNN however it was trained to reproduce the same data at the output as the input. Figure 4 illustrates the AANN reconstruction step that takes the place of the system difference identification step in Figure 2.

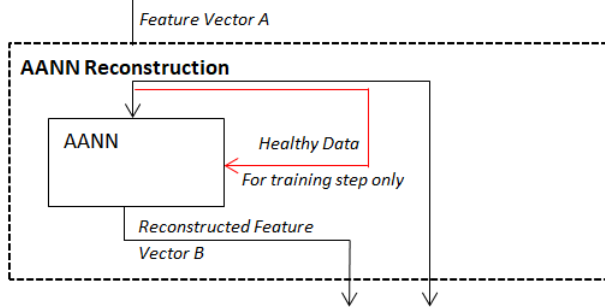


Figure 4. AANN reconstruction step

### Data Collection

Figure 5 shows the experimental apparatus used for data collection. Two ten-horsepower induction motors were used, one as a drive and one as a load. Both motors were controlled using variable frequency drive (VFD) motor controllers. The drive motor used a closed loop speed control and the load motor was set up for torque control. Between the two induction motors two gearboxes were connected in series one with a 3:1 ratio and the other 1:1.



Figure 5. Experimental Setup

The gearboxes were connected via a shaft with two Hooke's joints to accommodate misalignment. Figure 6, provides a schematic of the layout for the mechanical components in which the two gearboxes represent the quasi-parallel subsystems as described earlier. The gearbox speed and loads are directly related by the gear ratio of the gear-train but not identical. Further complexity to the system was also introduced by employing gears of different pitch in each gearbox resulting in different meshing frequencies (even for

same speed), stiffness and backlash properties. This leads to greatly different raw vibration signals from each gearbox even though they share common forcing functions.

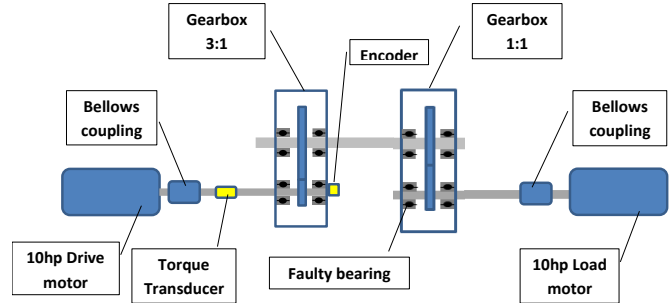


Figure 6. Arrangement of Mechanical Components

The machine was run on a duty cycle that includes three different steady state levels of speed and load as well as several different run-up and run-down conditions. The complete duty cycle can be seen in Figure 7.

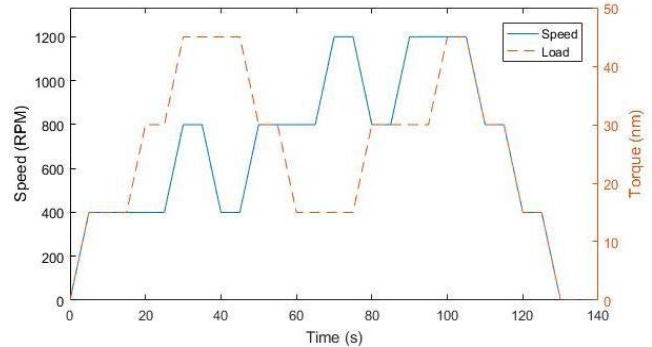


Figure 7. Duty cycle with independently controlled speed and load

The experimental data was collected from the machine using piezoelectric accelerometers (1mV/g sensitivity) and sampled at 10 kHz. Figure 8 shows the time series vibration signals from the two gearboxes for one run through the given duty cycle. In this case, both gearboxes are healthy. However, the difference between the two signals can clearly be seen. Moreover, it should also be noted that the signals in figure 8 also roughly illustrate the relationship between the vibrations in each gearbox as their changes in amplitude follow the same general pattern.

The accelerometers were mounted on the case of each gearbox, which is shown in Figure 9. The bearing trade number was 7616. Several different fault conditions were introduced into the bearings. These conditions were; rolling-element surface defect, outer race crack, inner race crack as well as multiple combined faults. In total, seven healthy and eight faulted tests were conducted (where each test constitutes one run through the duty cycle). These faults are described in Table 1.

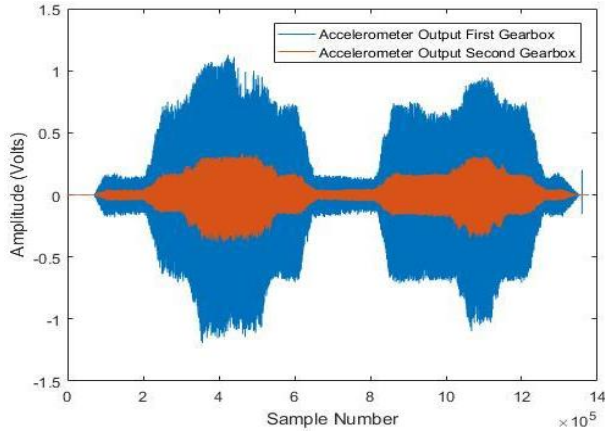


Figure 8. Time Domain Vibration Signals from Quasi-Parallel Gearboxes



Figure 9. Accelerometer and Encoder Mounted on Gearbox

Fault type	Label (For results section)	Description
Healthy	H(1-4)	No fault present, data used for training
Healthy	H(5)	No fault present, used for validation to stop overfitting during training
Healthy	H(6-7)	No fault present, data left for testing
Inner race crack	IR(1-2)	1-1.5mm in the inner race of the bearing
Outer race crack	OR(1-2)	Cut through the outer race of the bearing
Rolling-element surface defect	B(1-2)	1-1.5mm fault on one of the balls in the bearing
Combination of multiple faults	C(1-2)	1-1.5mm fault in outer race, inner race and rolling-element

Table 1: Fault Types and data set description

### 3. RESULTS

The results presented here illustrate the residual value for collected data segments. The residual value is taken as the normalized Euclidean distance between the network output and the second gearbox. Figure 10 shows a histogram of all residual values for all of the healthy and faulted data using the method described earlier. These results show significant separation between the two classes of data however 100% separation was not achieved as there is some overlap between classes. These results can then be compared to figures 11 to 13. Figure 11 provides the baseline comparison to a typical AANN. Whereas Figures 13 and 14 are included to further demonstrate to the reader the added value of including the neural network in the architecture and the raw features can't simply be compared to determine the health state of the system.

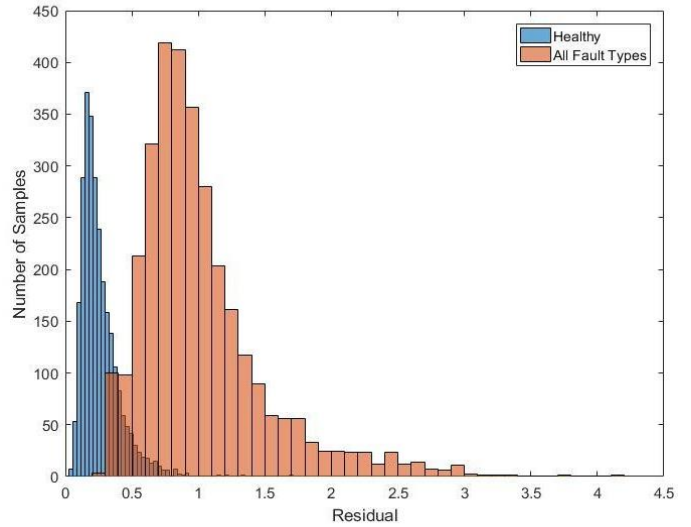


Figure 10. Histograms of the Residual Values for all of the Healthy and Faulted Data

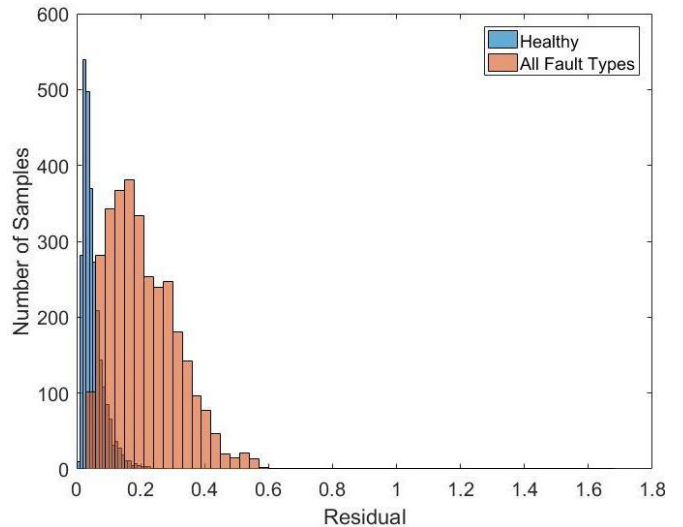


Figure 11: Residual Value with Non-Parallel AANN

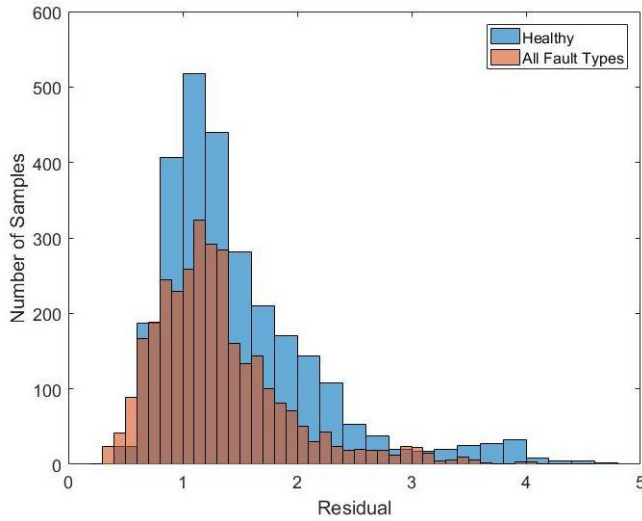


Figure 12. Histogram Parallel Residual Analysis Method (Without using an FFNN for system difference identification)

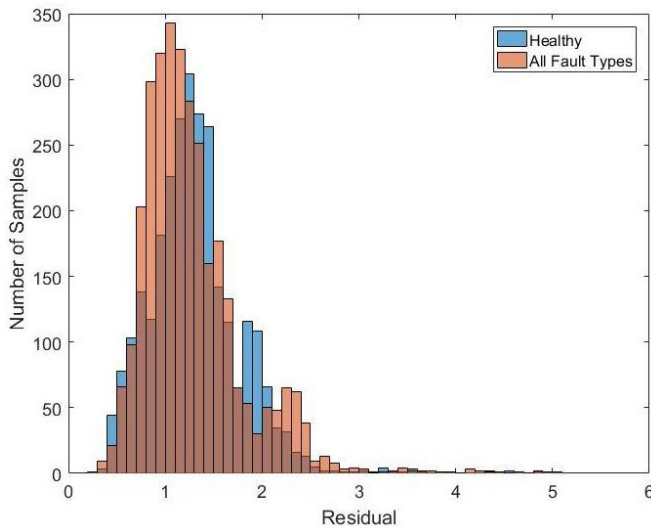


Figure 13. Histogram Non-Parallel Method (direct feature domain comparison to historical average)

Figure 11 shows a histogram for the same data processed in a non-parallel manner. The data is processed using an AANN novelty detector as described earlier. The network had the same structure and training methods as was used with the parallel method however the data was taken from a single gearbox with the network trying to reproduce the inputs at the output. This method utilized the same features and residual score calculation as in the parallel technique. There is still separation between the data classes however there appears to be a little more overlap which represents increased potential for incorrect classifications. This is further quantified in the classification results presented later. This loss in performance is attributed to the loss in

information due to analyzing the affected gearbox in isolation.

The results for the same data when processed without the use of the neural network can be seen in Figure 12 and Figure 13. Figure 12 shows the histogram of the data when analyzing the residuals between the two subsystems without the use of the neural network using the method presented in (D. Helm & Timusk, 2017). Figure 13 illustrates the results where the residual value is taken to be the difference between the feature vector and the average healthy value for the subsystem (non-parallel). These results show no separation between healthy and faulted tests. This demonstrates the improved accuracy resulting from adding the neural network to account for the relationship between the two subsystems.

To further quantify the effect of utilizing the proposed technique on classification results, the receiver operating characteristic (ROC) curve was generated for both the parallel FFNN and the non-parallel AANN methodologies. The ROC curve presents the fraction of healthy segments that lie below a threshold (targets accepted) versus the fraction of faulted segments have a residual score above the same threshold (outliers rejected). The ROC curves clearly show that the FFNN outperforms the AANN, see Figure 14. The area under the FFNN curve is 0.9836 and the area under the AANN curve is 0.9588.

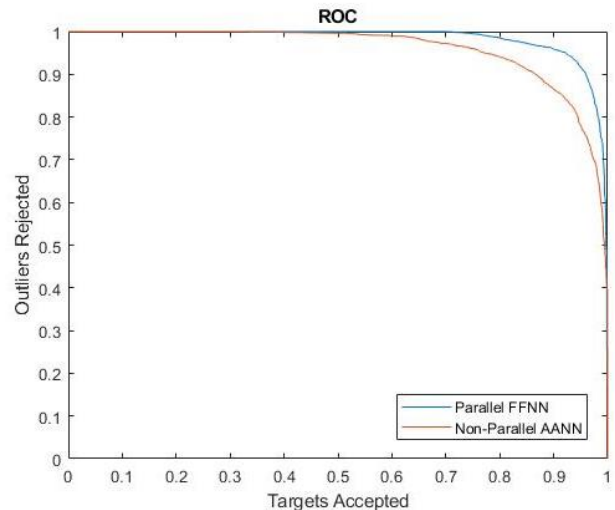


Figure 14. Receiver Operating Characteristic Curves

A thresholding classification method was applied to the residual values for each individual test. These results can be seen for the proposed parallel method as well as a standard AANN applied to a single gearbox in Figure 15. The threshold was set to exclude all but the top ten percent of the healthy training data. The results for the faulted data show greater than 90 percent accuracy for all the faulted tests with marked improvement for the parallel method over the AANN. These results are simply a snapshot of a single point



on the ROC curves in figure 14. In a real industrial application this threshold would be set based on the actual application and the relative priority of avoiding false negatives or false positives.

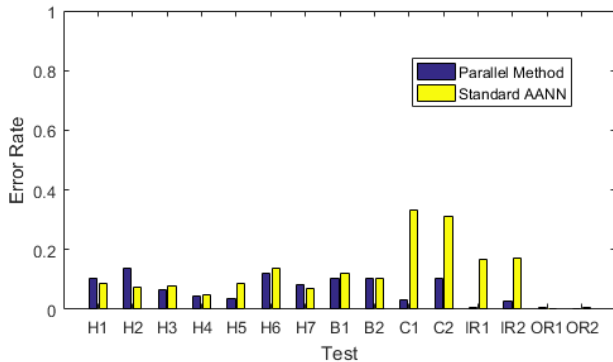


Figure 15. Thresholding Results for Parallel FFNN and non-parallel AANN

Another factor that was found to greatly influence the error rate is the length of the segments used for training and testing the neural networks. In all the previous results that length was set at 8 shaft rotations. The average error rate (over all the tests) can be seen to generally decrease with respect to the segment length in Figure 16. This may be expected as with longer segments there is less data and the results approach an average result over the entire test. While some of the segments will cover the short stationary parts of the duty cycle, others will contain variations in speed and load that will increase with the length of the segment. This increased variation does not appear to have negatively affected the results. However, there is a drawback to this as the system may not respond as fast to quickly changing operating conditions that are present in some machinery.

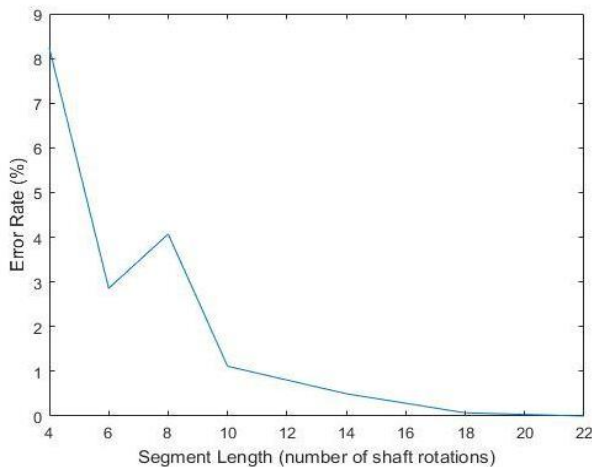


Figure 16. Segment Length vs. Error Rate

#### 4. CONCLUSION

It has been demonstrated that by incorporating a neural network to identify the relationship being subsystems it is

possible to extend experimental feature residual analysis as a technique for fault detection to systems that are not strictly parallel. It was also shown that in the case treated in this work by using a network similar to an AANN in conjunction with residual analysis, the detection accuracy of the condition monitoring system can be increased when compared to using an AANN directly as a novelty detector. The proposed method was demonstrated to be able to detect all of the fault types investigated here and provided a significant increase in accuracy for the inner race and combination fault compared to an AANN setup trained using the same data. These results show that by incorporating the data from parallel components into the fault detection scheme the error rate can be reduced. This is due to the forced relationship between the time varying parameters of the two subsystems allowing the fault detection system to remain insensitive to changes in the machine's operational conditions. While the potential benefits of the presented fault detection architecture have been demonstrated, it should be acknowledged that this method has some drawbacks. The proposed method is limited in possible applications and requires an extra signal channel that may not be present. Furthermore, this method is limited only to fault detection and does not diagnose the type or location of the fault. This method also requires the generation of features from the raw time-domain signal. Further work in this area could look into the incorporation of deep learning, however the tradeoff between complexity, amount of training data required, and performance should be closely examined. Another potential avenue for improvement could be optimizing the network architecture using a genetic algorithm or the application of different feature extraction techniques that have shown good performance for non-stationary systems, such as time-frequency domain analysis or cyclo-non-stationary indicators.

#### REFERENCES

- Abboud, D., Antoni, J., Sieg-Zieba, S., & Eltabach, M. (2017). Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment. *Mechanical Systems and Signal Processing*, 84, 200–226. <https://doi.org/10.1016/j.ymssp.2016.06.033>
- Abboud, D., Baudin, S., Antoni, J., Remond, D., Eltabach, M., & Sauvage, O. (2016). The spectral analysis of cyclo-non-stationary signals. *Mechanical Systems and Signal Processing*, 75, 280–300. <https://doi.org/10.1016/j.ymssp.2015.09.034>
- Barakat, M., Druaux, F., Lefebvre, D., Khalil, M., & Mustapha, O. (2011). Self adaptive growing neural network classifier for faults detection and diagnosis. *Neurocomputing*, 74(18), 3865–3876. <https://doi.org/10.1016/j.neucom.2011.08.001>
- Bin, G. F., Gao, J. J., Li, X. J., & Dhillon, B. S. (2012).

- Early fault diagnosis of rotating machinery based on wavelet packets - Empirical mode decomposition feature extraction and neural network. *Mechanical Systems and Signal Processing*, 27(1), 696–711. <https://doi.org/10.1016/j.ymssp.2011.08.002>
- Chen, Z., & Mechefske, C. K. (2001). Model order selection: A practical approach. *Mechanical Systems and Signal Processing*, 15(2), 265–273. <https://doi.org/10.1006/mssp.2000.1289>
- Cong, F., Chen, J., & Dong, G. (2012). Spectral kurtosis based on AR model for fault diagnosis and condition monitoring of rolling bearing. *Journal of Mechanical Science and Technology*, 26(2), 301–306. <https://doi.org/10.1007/s12206-011-1029-0>
- Figueiredo, E., Figueiras, J., Park, G., Farrar, C. R., & Worden, K. (2011). Influence of the autoregressive model order on damage detection. *Computer-Aided Civil and Infrastructure Engineering*, 26(3), 225–238. <https://doi.org/10.1111/j.1467-8667.2010.00685.x>
- Gertler, J. (1997). Fault detection and isolation using parity relations. *Control Engineering Practice*, 5(5), 653–661. [https://doi.org/10.1016/S0967-0661\(97\)00047-6](https://doi.org/10.1016/S0967-0661(97)00047-6)
- Gianluca, N., Fromageat, L., & Etienne, L. (2016). Machine Learning Strategy for Fault Classification Using Only Nominal Data. In *European Conference Of The Prognostics And Health Management Society*.
- Gor, M. M., Pathak, P. M., Samantaray, A. K., Yang, J.-M., & Kwak, S. W. (2018). Fault accommodation in compliant quadruped robot through a moving appendage mechanism. *Mechanism and Machine Theory*, 121, 228–244. <https://doi.org/10.1016/j.mechmachtheory.2017.10.011>
- Hagan, M. T., & Menhaj, M. B. (1994). Training feed-forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6), 989–993.
- Haidong, S., Hongkai, J., Xingqiu, L., & Shuaipeng, W. (2018). Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine. *Knowledge-Based Systems*, 140, 1–14. <https://doi.org/10.1016/j.knosys.2017.10.024>
- Hao, Y., & Wilamowski, B. M. (2011). Levenberg-marquardt training. In *Industrial electronics handbook* (5th ed., pp. 12-1-12–15).
- Helm, D. M., Rose, A. M., & Timusk, M. (2016). Condition monitoring of rolling-element bearings in parallel operating belt drive systems. *International Journal of COMADEM*, 19(3), 61–64.
- Helm, D., & Timusk, M. (2017). Using residual analysis for detection of faults in unsteadily operating rolling element bearings. In *WCCM 2017 - 1st World Congress on Condition Monitoring 2017*.
- Helm, D., & Timusk, M. (2019). Fault detection for parallel operating machines. *Journal of Quality in Maintenance Engineering*, 26(2), 335–348.
- Isermann, R., & Ball, P. (1997). Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 5(5), 709–719. [https://doi.org/10.1016/S0967-0661\(97\)00053-1](https://doi.org/10.1016/S0967-0661(97)00053-1)
- Japkowicz, N., Myers, C., & Gluck, M. (1995). A Novelty Detection Approach to Classification. In *14th International Joint Conference on Artificial Intelligence* (pp. 518–523).
- Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, 72–73, 303–315. <https://doi.org/10.1016/j.ymssp.2015.10.025>
- Jiang, H., Wang, F., Shao, H., & Zhang, H. (2017). Rolling bearing fault identification using multilayer deep learning convolutional neural network. *Journal of Vibroengineering*, 19(1), 138–149. <https://doi.org/10.21595/jve.2016.16939>
- Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 108, 33–47. <https://doi.org/10.1016/j.ymssp.2018.02.016>
- Kramer, M.A. (1992). Autoassociative neural networks. *Computers and Chemical Engineering*, 16(4), 313–328. [https://doi.org/10.1016/0098-1354\(92\)80051-A](https://doi.org/10.1016/0098-1354(92)80051-A)
- Kramer, Mark A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37(2), 233–243. Retrieved from [https://www.engineeringvillage.com/share/document.url?mid=cpx\\_1ef9f1d11c0a0637c629f62061377551&database=cpx](https://www.engineeringvillage.com/share/document.url?mid=cpx_1ef9f1d11c0a0637c629f62061377551&database=cpx)
- Lei, Y., Lin, J., He, Z., & Zuo, M. J. (2013). A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, 35(1–2), 108–126. <https://doi.org/10.1016/j.ymssp.2012.09.015>
- Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 108, 33–47. <https://doi.org/10.1016/j.ymssp.2018.02.016>
- McBain, J., Lakanen, G., & Timusk, M. (2013). Vibration- and acoustic-emissions based novelty detection of fretted bearings. *Journal of Quality in Maintenance Engineering*, 19(2), 181–198.
- Principi, E., Rossetti, D., Squartini, S., & Piazza, F. (2019). Unsupervised electric motor fault detection by using deep autoencoders. *IEEE/CAA Journal of Automatica Sinica*, 6(2), 441–451. <https://doi.org/10.1109/JAS.2019.1911393>
- Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics-A tutorial. *Mechanical Systems and Signal Processing*, 25(2), 485–520. <https://doi.org/10.1016/j.ymssp.2010.07.017>

- Rose, A. M., Helm, D. M., & Timusk, M. (2016). Fault detection of parallel hydraulic pumps in non-stationary operation. *International Journal of COMADEM*, 19(3), 55–59.
- Samanta, B. (2004). Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical Systems and Signal Processing*, 18(3), 625–644. [https://doi.org/10.1016/S0888-3270\(03\)00020-7](https://doi.org/10.1016/S0888-3270(03)00020-7)
- Samanta, B., & Al-Balushi, K. R. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical Systems and Signal Processing*, 17(2), 317–328. <https://doi.org/10.1006/mssp.2001.1462>
- Sanz, J., Perera, R., & Huerta, C. (2007). Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms. *Journal of Sound and Vibration*, 302(4–5), 981–999. <https://doi.org/10.1016/j.jsv.2007.01.006>
- Shao, H., Jiang, H., Zhao, H., & Wang, F. (2017). A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 95, 187–204. <https://doi.org/10.1016/j.ymsp.2017.03.034>
- Staroswiecki, M., & Comtet-Varga, G. (2001). Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems. *Automatica*, 37(5), 687–699. [https://doi.org/10.1016/S0005-1098\(01\)00005-X](https://doi.org/10.1016/S0005-1098(01)00005-X)
- Strczkiewicz, M., & Barszcz, T. (2016). Application of artificial neural network for damage detection in planetary gearbox of wind turbine. *Shock and Vibration*, 2016. <https://doi.org/10.1155/2016/4086324>
- Timusk, M., Lipsett, M., & Mechevske, C. K. (2008). Fault detection using transient machine signals. *Mechanical Systems and Signal Processing*, 22(7), 1724–1749. <https://doi.org/10.1016/j.ymsp.2008.01.013>
- Wang, G., & Cui, Y. (2013). On line tool wear monitoring based on auto associative neural network. *Journal of Intelligent Manufacturing*, 24(6), 1085–1094. <https://doi.org/10.1007/s10845-012-0636-7>
- Wang, X., & Makis, V. (2009). Autoregressive model-based gear shaft fault diagnosis using the Kolmogorov-Smirnov test. *Journal of Sound and Vibration*, 327(3–5), 413–423. <https://doi.org/10.1016/j.jsv.2009.07.004>
- Willersrud, A., Blanke, M., & Imsland, L. (2015). Incident detection and isolation in drilling using analytical redundancy relations. *Control Engineering Practice*, 41, 1–12. <https://doi.org/10.1016/j.conengprac.2015.03.010>
- Xie, Y., & Zhang, T. (2017). Fault Diagnosis for Rotating Machinery Based on Convolutional Neural Network and Empirical Mode Decomposition. *Shock and Vibration*, 2017, 1–12. <https://doi.org/10.1155/2017/3084197>
- Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 115, 213–237. <https://doi.org/10.1016/j.ymsp.2018.05.050>

## BIOGRAPHIES

**Dustin Helm** received his B.Eng (Bachelor of Engineering) degree in mechatronics engineering from Laurentian university, Sudbury, Ontario, Canada, in 2014. He is currently a PhD candidate at Laurentian university. His research interests are in the development of condition monitoring, and fault detection systems for non-stationary machinery.

**Markus Timusk** is a professional engineer and professor of mechanical engineering in the Bharti School of Engineering at Laurentian University, Sudbury, Canada. Dr. Timusk holds a PhD in Mechanical Engineering from Queen's University, Kingston, Ontario, Canada. He also holds Master's in Engineering Science from Western University, London, Ontario Canada. His undergraduate degree in Mechanical and Materials Engineering is also from Western University. His research interests involve, the development of decision support systems for the fault detection of mechanical systems including mobile mining equipment, industrial machinery and prototype automotive equipment. Prior to entering academia, he worked in the automotive industry developing vibration control devices for automotive engines and designing machinery to simulate machinery duty and vibration.