Towards Enhancing Maritime Asset Management Value Through Transforming Maritime Expert Knowledge into Machine Learning Models

[1a] Mohammad S. Abbas (MSc, BSc., MIMarEST, MRINA, CEng)

[2b] Ahmed A. Swidan* (Ph.D, FIMarEST, MRINA, FIEAust, CMarEng, CEng)

[a] Australian Shipbuilding Company, Henderson, 6166, WA, Australia

[b] Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt

Email*: Ahmed.Swidan@aast.edu.au

Abstract

Over the past decade, Machine Learning (ML) techniques have been utilised as an efficient decision making tool when big data is provided from systems' operators and/ or business developers. To generalise the diagnostic and prognostic analysis respectively for rolling element bearings, two ML models, namely; Weighted KNN classification and Linear regression were developed and are discussed in this paper. These models showed discrepancies in their performance when compared with domain expert's defined and models' defined training and testing datasets. However, considering the limited amount of training data used to train Weighted KNN model and the classification accuracy so obtained, it has been established that there is a strong potential of fostering expert knowledge into ML models. The presented approach can be employed effectively not only for a data-driven decision making to balance cost, risk and performance in a maritime asset management system but also for a better human training.

Keywords

Machine Learning; Maritime Asset Management; Condition Monitoring; Condition Indicators; Domain Expert.

1 Introduction

1.1 Background

Broadly defined, 'Asset Management' is the 'coordinated activity of an organisation to realise value from assets' (Standards Australia Limited, 2014). Asset Management philosophy encompasses all asset types, tangible and intangible, individual components or complex systems, and all activities involved in the asset's life cycle – everything from initial identification of requirements or opportunities, acquisition/creation, operations or utilisation activities, asset stewardship or care/maintenance responsibilities, through to renewal or disposal and any remaining liabilities (Global Forum on Maintenance and Asset Managment (GFMAM), 2014). This means Asset Management has a vast scope, and rather than just covering the maintenance of physical assets, its ultimate objective is to realise value from assets throughout the asset life. The values, which drive the Asset Management System of a defence maritime sustainment organisation, could be Capability, Availability and Affordability of the materiel they are sustaining.

A well-established condition monitoring (CM) program is emerging to be the core of all risk management based maintenance philosophies such as Condition-Based Maintenance (CBM), Reliability Centred Maintenance (RCM), Predictive Maintenance (PdM), Prognostics and Health Management (PHM) etc. (Cullum, Binns, Lonsdale, Abbassi, & Garaniya, 2018), and this is the point where these maintenance philosophies leap into the world of "Big Data' within the asset management framework. Big Data (Oracle Corporation, 2018) can be defined as the "data sets which are so voluminous that traditional data processing software just can't manage them". ML is one of the most popular techniques used for Big Data management, which apart from extracting useful information also drives the AI for a large number of engineering and other business applications (Han Jiawei, Chapter 2 - Getting to Know Your Data, 2012).

The Maritime Industry poses unique challenges from data acquisition, data quality and data accessibility point of view, in order to establish required condition monitoring analytics. CBM and PdM philosophies are getting acquaintance for maritime applications, but Periodic Preventive Maintenance (PPM) is still favoured by Maritime Industry due to the ease of resource planning (Cullum, Binns, Lonsdale, Abbassi, & Garaniya, 2018). In addition to this, due to the specific nature of missions that maritime materiel is supposed to undertake throughout its lifecycle, the probability of complying with the data acquisition schedule as per the approved condition monitoring programme is challenging and unlikely.

A key challenge in collecting high-quality data is the competence of shipboard crew which has the potential to directly affect the quality of measured data due to their lack of experience within this context. The data accessibility is also inhibited by security requirements for data protection, which may result in delays in accessing the data and restrict the amount of published data. In view of these challenges, ML provides a promising opportunity where the available condition monitoring data can be generalised to define the behaviour of a particular system, equipment or its component(s) with a limited variety of data, which may be just enough to define the condition indicators (CIs) required for a ML algorithm.

Furthermore, even in the absence of any design flaw, every machine operates in a unique system environment and owes to a set of specific inherent operational problems due to that particular system environment. Over a certain period of time, its operators and maintainers learn about operational issues of the machinery and eventually become 'domain experts' of the machinery for its various aspects, such as operation and maintenance etc.

Apart from the owners of the machinery, original equipment manufacturers (OEMs) of the machinery make best use of such domain expertise to improve their design and to develop services and products for their customers. In many instances when there is a consistent involvement of the operator and/or maintainer organisation with specific systems and its equipment, the domain expertise can reach to the level where they can make sometimes better decisions than OEM for the operation and maintenance of their system and equipment within a particular system environment. A classic example of domain expertise's precedence over OEM is the merchant shipping industry where the direct involvement of OEM, outside the warranty obligations, is considered to be an expensive solution and all the technical decisions are made by the ship crew and shore management balancing vessel's contractual engagement and commercial viability. ML in combination with artificial intelligence (AI) has the capability to foster such domain expertise particularly for risk management based maintenance philosophies. This is the main driving motivation to undertake this industrial research project which is mainly focused on identifying and addressing the challenges, risks and opportunities of transforming expert knowledge into a ML model.

Such a ML model has a potential to provide a data-driven framework in the maritime asset management system, which can be employed effectively to balance cost, risk and performance in the value (capability, availability and affordability) realisation of maritime assets.

1.2 Research Objectives

This research is aimed to explore solution propositions for the following notions:

i. How can we transform human expert knowledge about any process or analysis into ML models?

ii. How can we validate and improve the effectiveness of a ML model developed based on the available human expert knowledge?

iii. Can a human expert knowledge based ML model add value to an asset management system in maritime context?

2 Numerical Model

ML describes purposely designed "algorithms which use computational methods to 'learn' information directly from data without relying on a predetermined equation as a model" (MathWorks, Inc., 2016). The most interesting aspect of ML is that its algorithms improve their performance with an increase in the number samples used for learning. ML has emerged as the game changer in the space of data science and AI in the recent years (Press, 2015).

However, adopting ML poses its own challenges which include the requirement of using an enormous amount of data in order to generalise a ML model with a high level of accuracy, and with the limited availability of people with right skills to adopt this technology. Some (Marouani, 2018) believe that it is high time that engineers need to develop their skills in the domain of data science in general, and for ML development and its implementation in particular, in order to fill the gap of a lack of data scientists. Figure 1 provides mapping of commonly used ML techniques, methods and models, developed based on the learning analysis from (MathWorks, Inc., 2018), (MathWorks, Inc., 2016), and (Witten, 2005).

2.1 Developing a ML Model

Developing a ML model in a real world scenario is predominantly an iterative process. It involves test and trial of different ideas emerging from the developer's domain expertise in ML modelling and in the process to be generalised through ML modelling. Figure 2a illustrates typical workflow of a ML model development, established based on the learning analysis from (MathWorks, Inc., 2018) and (Han Jiawei, Chapter 2 - Getting to Know Your Data, 2012). Figure 2b illustrates the average time consumed to accomplish each stage of the ML model development workflow based on 0.125 equivalent full time study load (EFTSL).

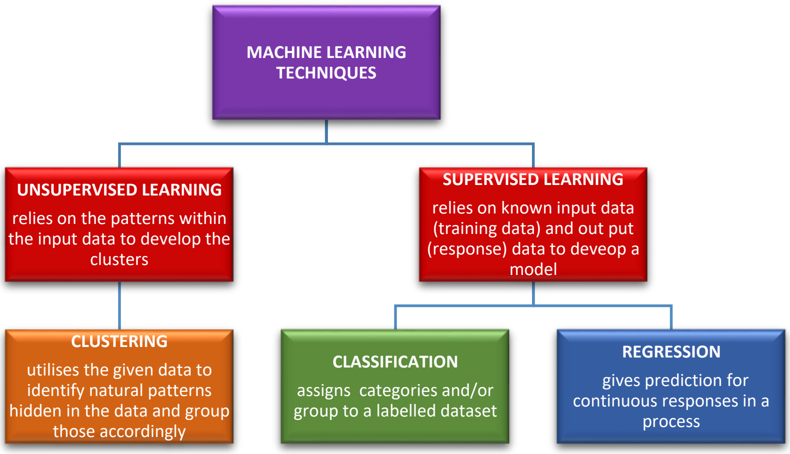


Figure 1a: Broader classification of Machine Learning techniques

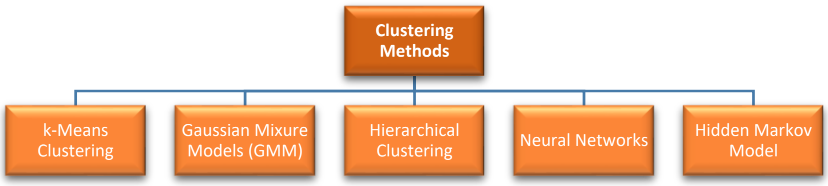


Figure 1b: Clustering methods flowing down from Figure 1a

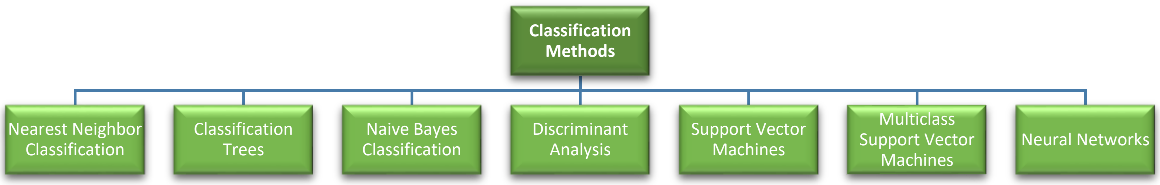


Figure 1c: Classification methods flowing down from Figure 1a

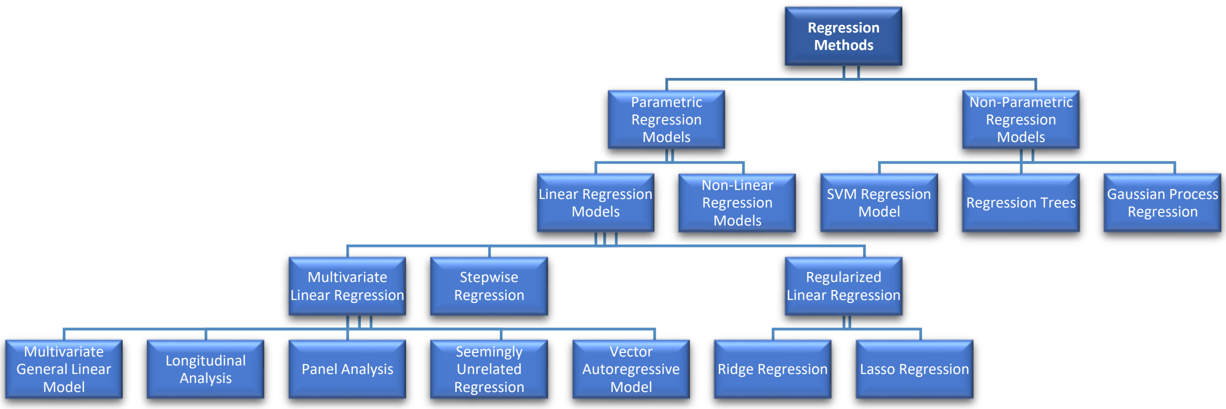


Figure 1d: Regression methods flowing down from Figure 1a

Figure 1: Figure 1: Mapping of ML techniques, methods and models.

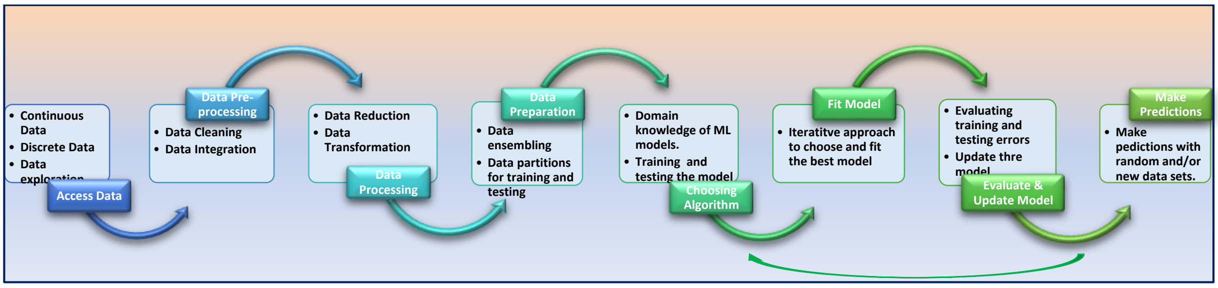


Figure 2a: Typical workflow of ML model development, based on the learning analysis from (MathWorks, Inc., 2018) and (Han Jiawei, 2012)

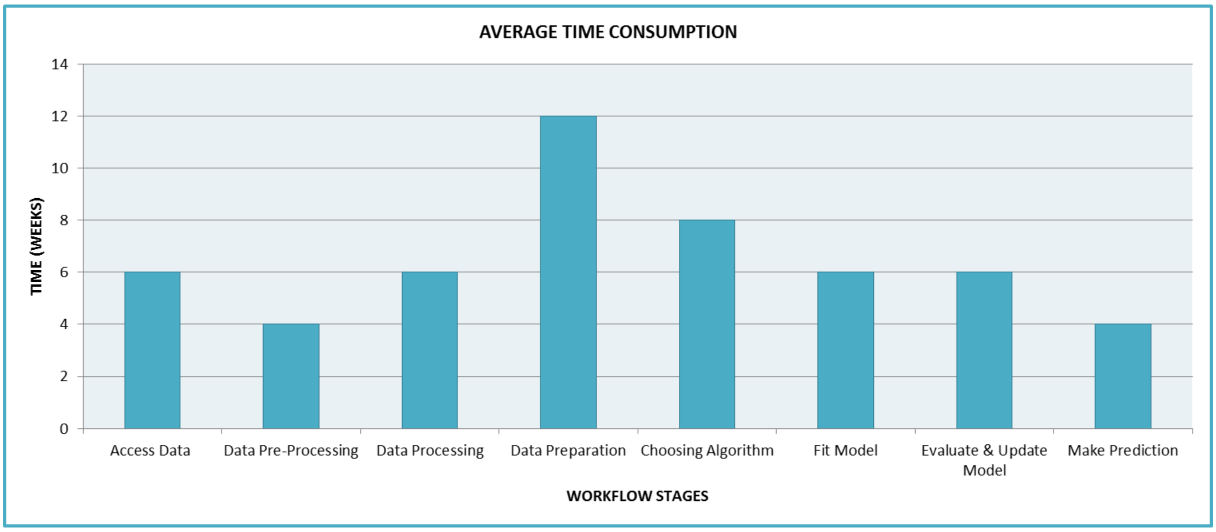


Figure 2b: Average time consumption for each stage of work flow based on 0.125 EFTSL

In order to develop a prototype ML model in this research, a commonly used and tested machinery component which was under the CM program was selected as the test subject. Vibration analysis datasets obtained from rolling element bearings of similar specification but used in various rotating equipment applications were chosen for the ML model development. The specifications of the selected rolling element bearing model and design cannot be disclosed due to commercial and other reasons.

Bearing defect frequencies, also known as fundamental fault frequencies, of a rolling elements bearing play an important role in its vibration analysis. These frequencies are the function of bearing geometry (i.e. pitch diameter and roller diameter), the relative speed between the two raceways and the number of rolling elements (Rolling Element Bearing Analysis, 2012). For a known bearing geometry, the set of fundamental fault frequencies can be calculated as illustrated in Figure 3. This research works with the relative amplitudes of the fundamental fault frequencies for rolling element bearing of chosen specification to develop a ML model.

A lot of theoretical work has been conducted previously in relation to the application of ML for fault analysis of rolling element bearings which predominantly discusses how various techniques of ML can be applied for bearing fault analysis, classification and training, such as: (A Comparative Study on Bearings Faults Classification by Artificial Neural Networks and Self-Organizing Maps using Wavelets, 2010), (Ball Bearing Fault Diagnosis Using Supervised and Unsupervised Machine Learning Methods, 2015), (Fault diagnosis of rolling element bearings via discriminative subspace learning: Visualization and classification, 2014) and (Remaining Useful Life Prediction of Rolling Bearings Using PSR, JADE, and Extreme Learning Machine, 2016) are to name a few. However, this paper focuses on the application of most suitable ML techniques in a real world scenario using the algorithms and toolboxes of MATLAB in order to get the best possible generalisation of ML model such that it should identify different stages of bearing faults in its life cycle before reaching a permanent failure.

2.2 Data Gathering

The research made use of two datasets; a training dataset, and a testing dataset. The training dataset was comprised of 22 labelled samples of bearing vibration data and was used to train the ML model. The testing dataset was comprised of 12 random samples of bearing vibration data and was used to evaluate the generalisation of the developed model. Each data sample included 1600 observations of Amplitude (gSE-Peak), measured corresponding to each sampling frequency interval. Each of the labelled data samples (training data samples) included one of bearing fault stages i.e. 'Normal', 'Fault Stage 1', 'Fault Stage 2', and 'Fault Stage 3', according to the judgement of the domain experts. However, the testing data samples were without any label attached pertaining to any of the bearing fault stages.

Equation 1 – Ballpass frequency, Outer race (BPFO)

$$BPFO = \frac{nf_r}{2}\left(1 - \frac{d}{D}cos\emptyset\right)$$

Equation 2 – Ballpass frequency, Inner race (BPFI)

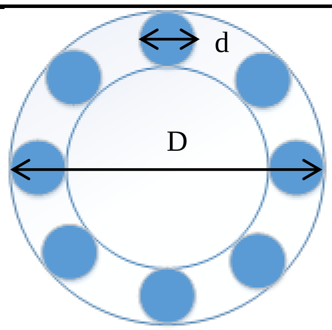
$$BPFI = \frac{nf_r}{2}\left(1 + \frac{d}{D}cos\emptyset\right)$$

Equation 3 – Fundamental Train Frequency (FTF)

$$BPFO = \frac{nf_r}{2}\left(1 - \frac{d}{D}cos\emptyset\right)$$

Equations 4 – Ball (roller) Spin frequency (BSF)

$$BSF = \frac{D}{2d}\left[1 - \left(\frac{d}{D}cos\emptyset\right)^2\right]$$



$d = Ball\ Diameter;\ D = Pitch\ Diameter$

$f_r = Shaft\ Speed;\ \emptyset = Bearing\ Contactg\ Angle$

$n = number\ of\ rolling\ elements$

Figure 3: Equations of fundamental fault frequencies (Rolling element bearing diagnostics—a tutorial, 2011)

Each sample also contained the manufacturer's defined ratios of BPFI, BPFO, BSF and FTF, as presented in Figure 3.

2.3 Data Pre-processing and Processing

Prior to reaching the stage of a ML model development, it takes an enormous effort to organise the data required to test a chosen model. Analysis (CrowdFlower, Inc., 2016) reveals that data scientists spend only 19% of their time in collecting the datasets, whereas they spend 60% of their time in cleaning and organising datasets before they can prepare the data for further application. Figure 4 illustrates the data pre-processing and data processing techniques and process based on the learning analysis from (Han Jiawei, Chapter 2 - Getting to Know Your Data, 2012). It reveals that data pre-processing and data processing comprise of four main processes as follows:

a. Data Cleaning – It involves applying the appropriate data science techniques to remove the noise and rectifying inconsistencies in the data (Figure 4a).
b. Data Integration – This implicates merging data from multiple sources into consolidated object such as data ensembles (Figure 4b).
c. Data Reduction – Data reduction techniques are applied to extract the data of interest and importance for intended application (Figure 4c).
d. Data Transformation – Data transformation techniques are extensively applied to make the data normalised, standardised and smoothened for further processing in the work flow (Figure 4d).

Although all the four processes have some common techniques which can be applied from one process to the other, a set of solid data science skills is required to make selection of the appropriate technique as similar techniques do not produce similar results for a given dataset due to their limitation based on the size of data, different boundary conditions and theory of application. An iterative approach prevails to get best possible results in data processing.

Preparing the data to progress through workflow (as illustrated in Figure 2a) using MATLAB required the datasets to be stored in a specific format. This was achieved by transforming the data into a specific format suitable for input to ML algorithms in MATLAB and then developing ensemble data store objects using MATLAB defined functions.

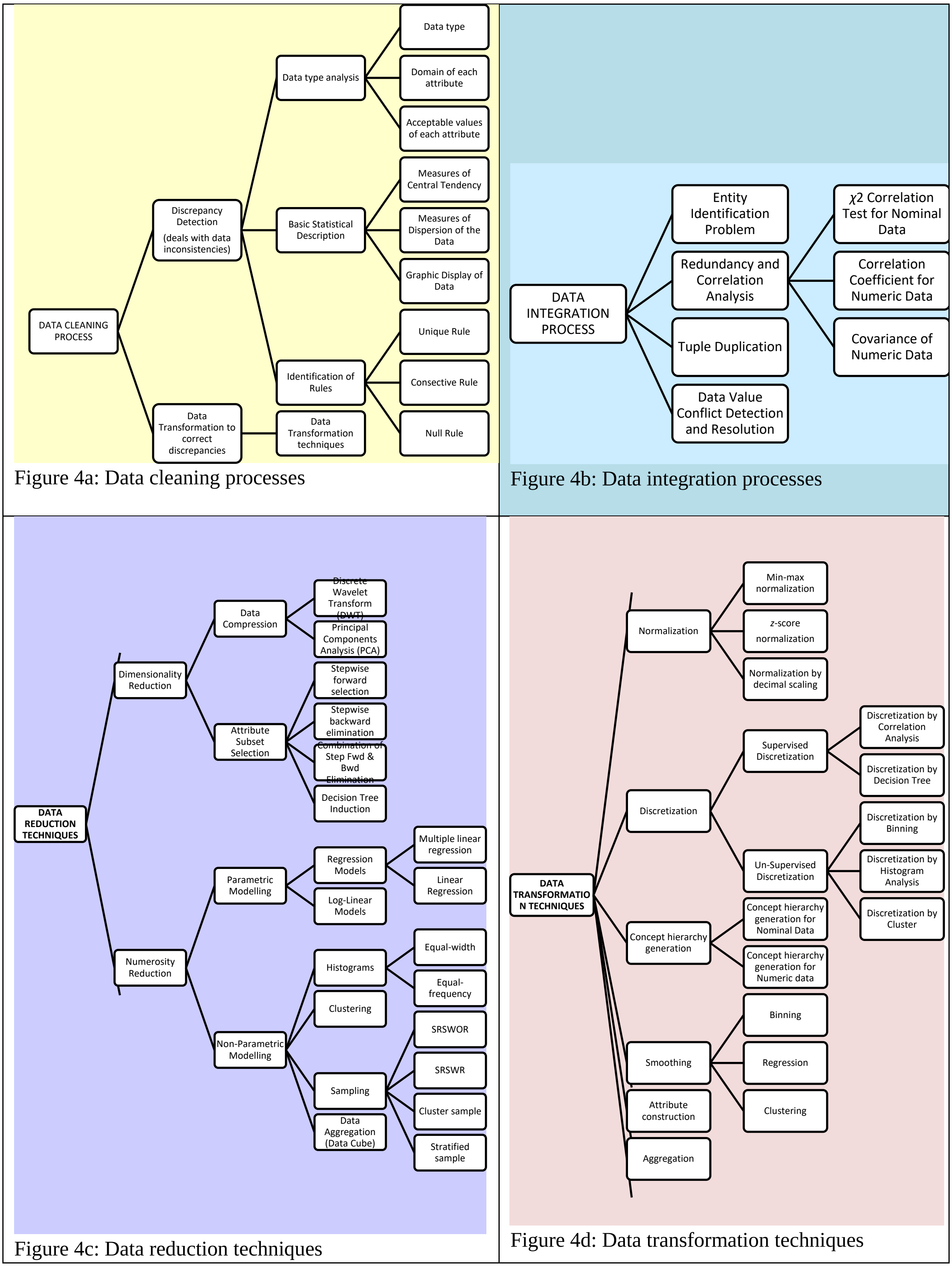


Figure 4a: Data cleaning processes

Figure 4b: Data integration processes

Figure 4c: Data reduction techniques

Figure 4d: Data transformation techniques

Figure 4: Data pre-processing (Figure 4a and 4b) and Data processing (Figure 4c and 4d) processes and techniques, based on the learning analysis from (Han Jiawei, 2012).

MATLAB has a number of specially defined functions which are used to manage relatively large and complex datasets of machinery operations and condition monitoring etc. by creating ensemble data store objects which can be further used for predictive maintenance and machine learning algorithms (MathWorks, 1994-2018). The two datasets (i.e. training and testing dataset) were transformed into separate ensemble data stores for training and testing of the ML model respectively prior to their application for ML algorithms.

3 Methodology

Figure 5 illustrates the methodology employed in a sequential manner to accomplish the research objectives.

(Zhou, et al., 2018) reveals the importance of bringing together the end users and developers of a ML model in order to build the confidence of end users about the capabilities, strengths and weaknesses of the developed ML model. This approach necessitates maximum engagement of the stakeholders (i.e. domain expert of vibration analysis in this case) from the concept development phase to product deployment. This approach has been incorporated in the methodology through feedback loops, as illustrated in Figure 5.

3.1 Selection of CIs

Bearing defect frequencies are the set of frequencies presented when rolling elements strike a localised fault on the inner or outer race, or conversely, a fault on any rolling element strike the inner or outer race (Bearing Envelope Analysis Window Selection, 2009).

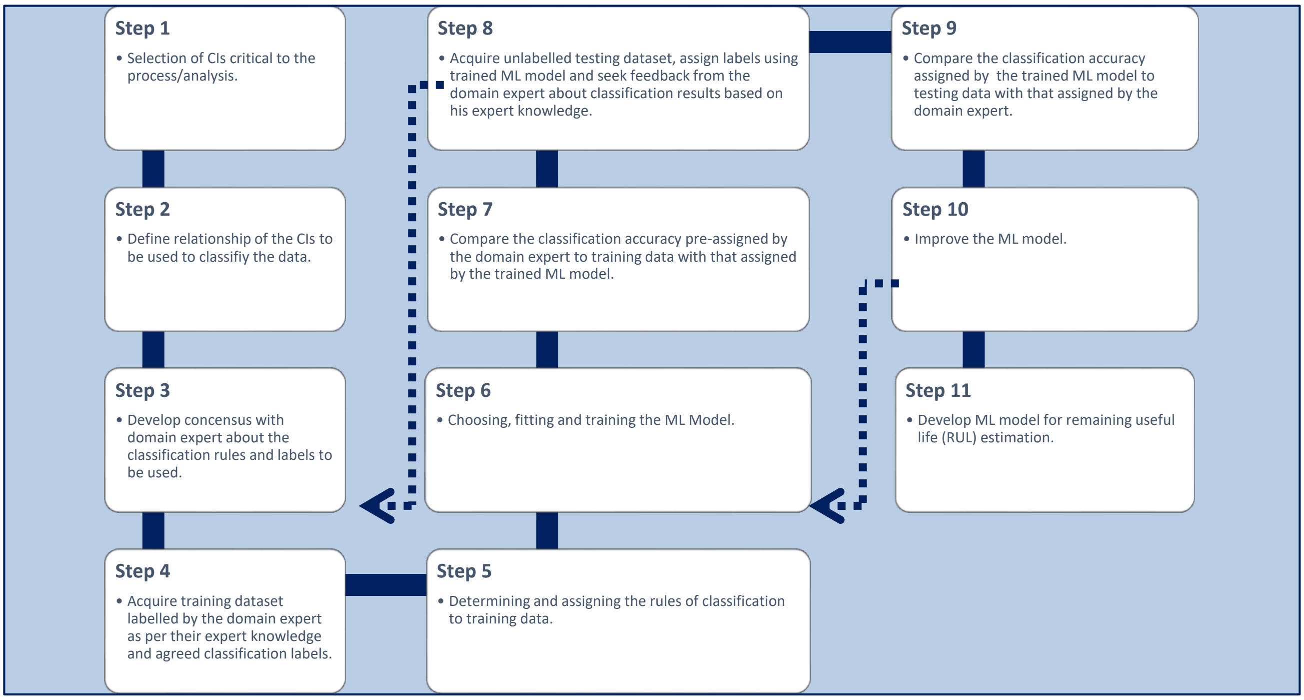


Figure 5: Work flow of methodology

The result of this phenomenon is the production of cyclic impacts which can be traced in bearing envelope analysis, due to an increase in vibrational energy at the time of impact. Analysis (An Overview on Vibration Analysis Techniques for the Diagnosis of Rolling Element Bearing Faults, 2013) reveals that the most common contributors towards the initiation of localised faults are cracks, pits and spalls and their cause usually attributed to a number of interlinked mechanisms such as excessive loads, fatigue, lack of lubrication and excessive temperatures etc. These faults have a strong likelihood of propagating with different operating conditions.

Identifying the deterioration trends of a rolling element bearing and attributing a right cause to each of these trends is not a straight forward process. Generally (Howieson, 2003) the harmonics appearing in a Fast Fourier Transform (FFT) of time domain signals of a defective bearing can be attributed to a combination of non-bearing related defects such as looseness, misalignment, or unbalance or equipment's internal defects etc. Such a combination of harmonics produces a 'haystack' effect in the high frequency region of the spectrum. However, it can be misleading if the frequency spectrum is analysed in isolation. Bearing envelope analysis, on the other hand, involves the demodulation of high frequency resonance associated with bearing element impacts (Bearing Envelope Analysis Window Selection, 2009). In other words, the impacts due to a localised fault modulate a vibration signal at the associated fundamental fault frequencies of the bearing. The vibration signal is then passed through a high pass (or band pass) filter which rectifies the signal into half or full wave. This rectified signal is passed further through a low pass filter which separates the modulation (or defect) frequency from carrier frequency.

Analysis (Bearing Envelope Analysis Window Selection, 2009) also reveals that the amplitudes associated with each of the bearing fault frequencies can be used as CIs reflecting the health of the bearing components. However, using four CIs corresponding to each fundamental fault frequency of the bearing, it is very likely that each sub element (i.e. ball, train, inner and outer race) of the bearing will have its own resonant mode. As a result, a selection of window for frequency and bandwidth to capture the amplitudes associated with one of the fault frequencies of the bearing might not be an optimal choice for other three sub elements. Moreover, as CI values indicate health in different parts of the spectrum, each will have different nominal amplitude values.

The load of the machine was not taken into consideration for spectrum analysis and amplitude calculations, in order to develop a prototype ML model with a set of minimum predictor variables.

3.2 Relationship of CIs to classify data

The ML model to be developed postulates the relative values of amplitudes associated with the BPFI and BPFO (see Section 1.3 and Figure 3) can be used as the CIs of the rolling element bearing if we use the log ratio of BPFI and BPFP amplitudes {i.e.log $\left(\frac{BPFIAmplitude}{BPFOAmplitude}\right)$} to categorise various fault stages in the lifecycle of the bearing (MathWorks, Inc., 1994-2018). With an understanding (Rolling Element Bearing Analysis, 2012) of the four fault stages in the lifecycle of a rolling element bearing, the lifecycle of the bearings used in this research are considered to be comprised of three fault stages (i.e. Fault Stage 1, Fault Stage 2, Fault Stage 3), avoiding Fault Stage 4 in order to prevent any 'run to failure' scenario in accordance with applicable maintenance policies.

Furthermore, the domain expert would also determine any particular 'Fault Stage' (assigning the label) in the training dataset based on his 'expert knowledge' and this classification would be compared with that assigned by the trained ML model as described in the Steps 4 to 7 of the methodology (Figure 5).

3.3 CI extraction

In order to extract two CIs as discussed above, a set of pre-existing MATLAB codes were modified to develop a batch processing algorithm meeting the requirements of available data specifications and the abovementioned ML model hypothesis. This algorithm makes use of MATLAB functions such as envelope spectrum analysis, spectrogram and spectral kurtosis, in order to extract the relative values of amplitudes associated with the BPFI and BPFO from each of the samples of given dataset (MathWorks, Inc., 1994-2018).

4 Results and discussion

Section 3 discusses in detail the developed methodology and the gained results. The section also includes a detailed uncertainty analysis to provide the reader with confidence in the computed results.

4.1 Determining and assigning the rules of classification to training data

This step involved the determination of log ratio of relative amplitudes of BPFI and BPFO {i.e.log $\left(\frac{BPFIAmplitude}{BPFOAmplitude}\right)$}in the training dataset and then assigning a best-fit range of these values to each of the four class labels i.e. 'Normal', 'Fault Stage 1', 'Fault Stage 2', and 'Fault Stage 3', as illustrated in Figure 6. A customised MATLAB code was used to extract the values of relative amplitudes associated with BPFI and BPFO and their log ratio for each of the given samples of training dataset stored in an ensemble data store. The log ratio vector of training data samples was discretised into four bins using the histogram function of MATLAB. The bin edges obtained approximately define the boundaries of each of the class label as mentioned in Figure 6a. Applying the histogram function on pre-labelled training data samples by the domain experts, a general trend for each of the classification label was revealed and the bins were assigned 'labels' as mentioned in Figure 6b and 6c. Figure 6d shows the sample counts in a narrower bin width in each of the assigned class. The training data samples were relabelled by applying a newly defined classification rule at their log ratio vector and results were then compared with the class labels assigned by the domain experts.

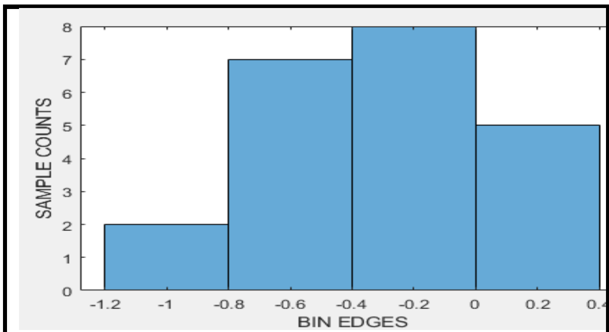


Figure 6a: Bin edges of log(BPFIAmplitude/BPFOAmplitude) defining the boundaries of each of the class label

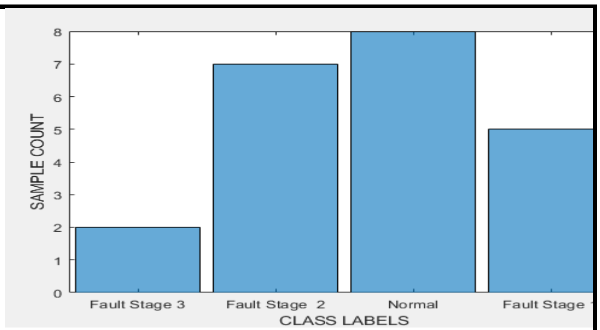


Figure 6b: Class labels corresponding to each bin

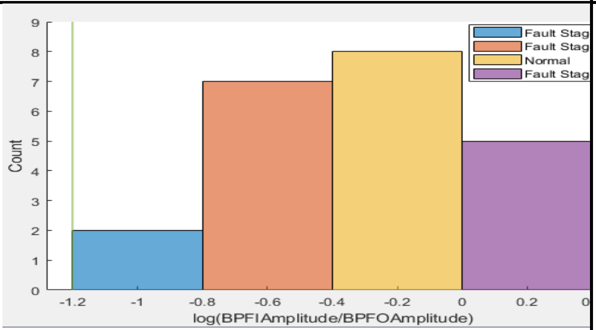


Figure 6c: Bin edges and class labels

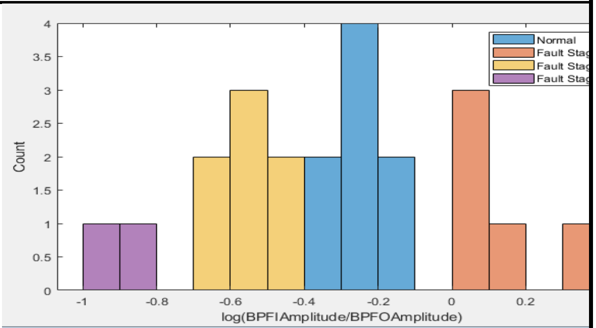


Figure 6d: Sample counts in each of the defined class bins

Figure 6: Determining and assigning the rules of classification

The classification accuracy of approximately 55% was recorded according to the defined classification rule as illustrated in Figure 6a. The error of 45% due to the misclassification of pre-labelled data (defined by the domain experts) could be attributed to the limited number of data samples, and also no distinctly and specifically defined boundary distinguishing one stage fault from another in the spectrum and envelope analysis of vibration signals, thus providing the domain expert a leverage of open choice to make judgement about the fault stage. It is, therefore, highly likely that human judgement can assign a misclassified label relative to the classification defined in Figure 6 as it can vary from person to person due to his/her expertise and experience in vibration analysis.

4.2 Choosing, fitting and training the ML Model

A specific data table was created in MATLAB for choosing, fitting and training the ML using MATLAB's Classification Learner App (CLA) (MathWorks, Inc., 1994-2018). The data table was comprised of three predictor variables i.e. 'BPFI Amplitude', 'BPFO Amplitude' and their log ratio {i.e.log $\left(\frac{BPFIAmplitude}{BPFOAmplitude}\right)$}; the response variable being 'Class Label' as discussed in the Section 3.1. The vector 'Class Label' was comprised of data labels assigned to each training data sample according to the determined classification rule as per the Figure 6. CLA in MATLAB allows using the training data for a number of classification models simultaneously with an output of overall validation accuracy score (expressed in percentage) displayed for the models trained. The validation accuracy score is the indicator of a model's performance on new data compared to the training data (classification performance). Making use of this capability of CLA, weighted nearest neighbour classifier (Weighted KNN) with an accuracy of 90.9% (as calculated by CLA) was finally selected for further analysis. A Weighted KNN model works on the principle of making distinctions between the classes using a distance weight (MathWorks, Inc., 1994-2019). The algorithm applied by CLA to develop Weighted KNN model categorised the training data into 10 numbers of neighbours based on the distance between the standardised data points and applied Euclidean metric for distance determination. The algorithm also used squared inverse method to determine the distance weight between the classes. Another reason for making the choice of Weighted KNN over other models with the similar accuracy scores was the simplicity and effectiveness of the model particularly given all the predictors of the training datasets are numeric.

4.3 Visualisation of ML Model

The appropriateness of chosen ML model can be further analysed and compared with other ML models trained with the same training data by using various visual techniques available in MATLAB and its CLA function. Such techniques include scatter plots, confusion matrix, parallel coordinates plot, Receiver Operating Characteristics (ROC) curve etc. (MathWorks, Inc., 1994-2019).

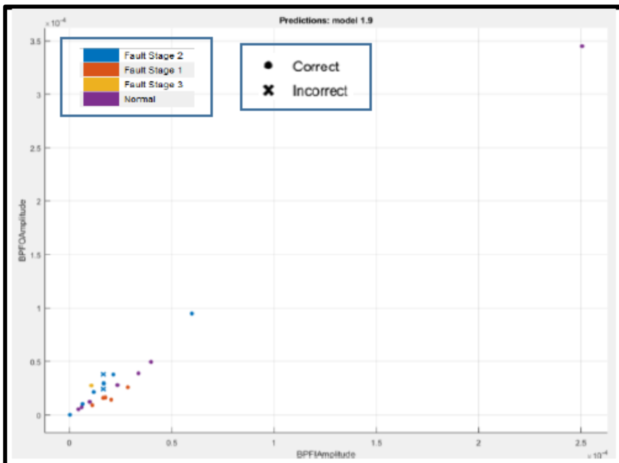


Figure 7a: Scatter plot of Weighted KNN model gives visualisation about the misclassified points out of the training data

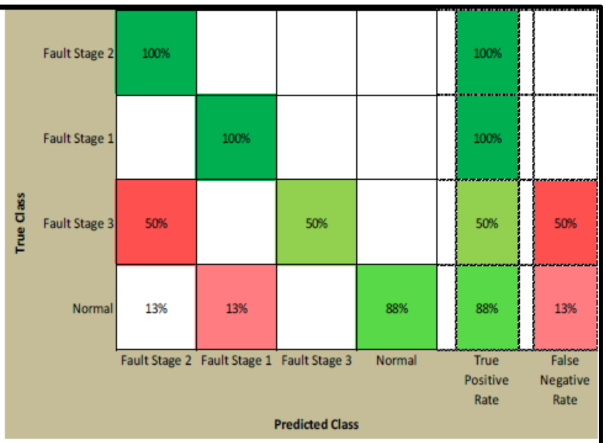


Figure 7b: Confusion matrix showing the numerical as well as heat map indications for the match between true and predicted classes of the Weighted KNN model

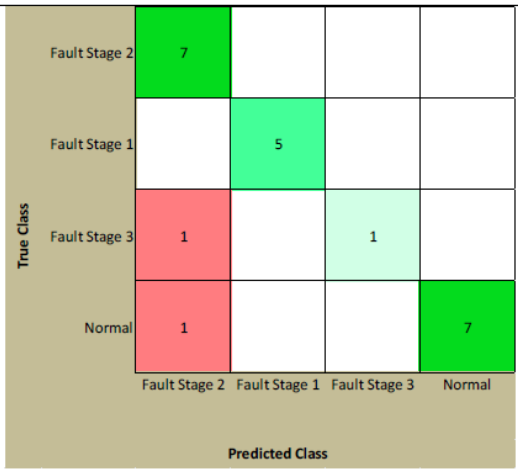


Figure 7c: Confusion matrix in terms of no. of observations

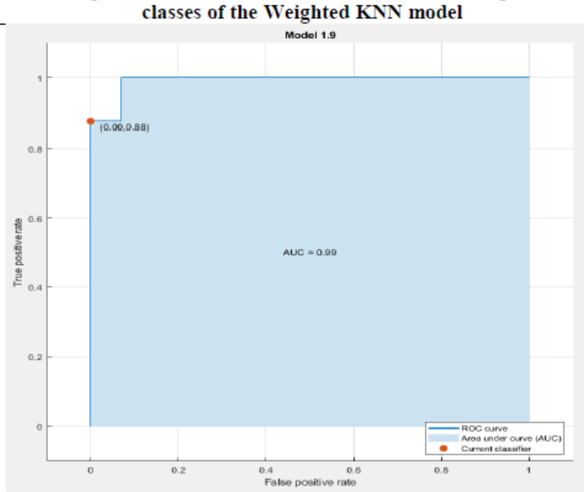


Figure 7d: ROC curve in terms of TPR and FPR

Figure 7: Visualisation of and performance of Weighted KNN model

The scatter plots give visualisation about the misclassified points out of the training data and also help to make feature selection in order to get best possible generalisation of chosen ML model. Figure 7a is the scatter plot of chosen ML model i.e. Weighted KNN and shows the misclassification of two points. Parallel coordinates plot is another technique commonly used for feature selection. However, due to very limited number of predictors, feature selection is not applicable in this case.

The model can be further analysed in confusion matrix plots which provide an indicative performance evaluation for each of the class in the ML Model, as mentioned in Figure 7b for Weighted KNN model. The diagonal cells in Figure 7b provide the numerical (in percentage) as well as heat map indications for the match between true and predicted classes, and True Positive Rate (TPR) and False Negative Rate (FNR) for each classifier. Figure 7c provides the similar visualisation in terms of actual number of observations for each of the class.

ROC curve is another technique to assess the performance of each of the classifier included in the chosen ML model in terms of TPR and false positive rate (FPR). Figure 7d shows the TPR of 0.88 for the classifier 'Normal', in concurrence with TPR mentioned in Figure 7b, which means that the class 'Normal' assigns 88% of the observations correctly in the chosen Weighted KNN model. Similarly the ROC for each of the class can be visualised separately.

4.2 Testing the trained ML model

Testing of the developed ML model involved making predictions of classes for new datasets. In other words, the developed ML should allocate a correct class label to each of the random samples of bearing vibration data in the Testing Dataset. The accuracy of the prediction is the measure of generalisation of the developed ML model. The testing dataset, comprised of 12 random (unlabelled) samples of bearing vibration data, was processed and prepared for ML model testing using once again a customised MATLAB code (as discussed in Section 1.3.2) which also extracted the values of three predictor variables i.e. relative amplitudes associated with BPFI and BPFO and their log ratio for each of the given samples of testing dataset stored in a separate ensemble data store.

The predict algorithm of Weighted KNN model was then run to get the class labels for each of the data sample. The model performed with 0 % testing error i.e. none of the data samples in testing dataset was misclassified by the ML model. The reason for this perfect performance can be attributed to the quality of testing dataset in which none of the samples had any parameter unrecognised by the trained Weighted KNN model. The test results for each of the samples were shared with domain experts and their concurrence with the model test results based on their expert knowledge was obtained. Comparing the concurrence results received from the domain experts with that of produced by the ML model, it was revealed that the ML model assigned the class labels to 42% of the samples correctly relative to the expert judgement i.e. 58% of the samples were not classified correctly as per the expert judgement.

It was also found that the misclassifications allocated by the domain experts were closely interfaced in terms of fault stages in the most cases.

4.3 Comparison of classification accuracy

The relative classification accuracy of training data (as discussed in Section 4.1) and testing data (as discussed in Section 4.3) with respect to both ML model and expert judgement results were validated and the summary of the comparison were presented in Figure: 8.

	Classification Accuracy for Training Data	Classification Accuracy for Testing Data
ML Model	55%	100%
Expert Judgement	100%	42%

Figure 8: Comparison between ML model and expert judgement for relative classification accuracy of training and testing data

Learning (MathWorks, Inc., 2018) also reveals that classification accuracy required from a ML model is driven by the importance of information in a given decision making framework. For example, the accuracy of information required by a ML model in a medical diagnostics framework cannot be compared with the one which is used for making decisions in a typical condition monitoring program. Additionally, the more complex a ML model is to fit the training data near perfection, less efficiently it would perform with the new datasets. However, in any case the generalisation and thus the accuracy of a ML are dependent on the amount and variety of training datasets. Apart from human error, the discrepancies in expert judgement with respect to the developed ML model can also be attributed to the system's environment specific operational features taken into consideration by the domain experts, which are not the part of training data.

In case of Weighted KNN model which is a prototype ML model developed for this research, the classification accuracy (as mentioned in Figure 8) appears reasonable considering the fact that the model was trained with one dataset containing only 22 data samples with four different class labels.

The most desirable way to minimise accuracy discrepancies between the human judgement and Weighted KNN model is an iterative training of the ML model with as many data samples for each of the class label as possible.

If the ML model performance does not improve proportionately to the iterative training, the other options worth considering could be reviewing the rules of classification such as reviewing the bin edges (as mentioned in Figure 6) so as to include the misclassified samples and/or fitting other ML models most appropriate to the data type, quality and quantity. Other contributing factors to maximise the accuracy of a ML model could be improving the human judgment with further training of domain experts and/or operators and incorporating system's environment specific dynamics in the training data.

4.4 Estimating Remaining Useful Life (RUL)

Estimating RUL is one of the main objectives of prognostic analysis for any system, equipment or its components. It generally involves the development of a model which can provide estimation of RUL based upon the time evolution or statistical properties of CI values (MathWorks, Inc., 1994-2019). Prognostic analysis can make use of mathematical models, machine learning or a combination of both to predict the values of CIs which can be used to compute RUL metrics (MathWorks, Inc., 1994-2019). However, in order to develop such RUL models with an acceptable accuracy a large amount of life data is required. In the case of this research, using expert knowledge about the average time a bearing takes to develop one of the labelled fault stages, and then the average time it requires to reach the stage where it should be considered for an immediate replacement, a simple ML regression model can be developed.

Given the amount of data and significant feedback from domain experts, a linear regression ML model was developed using 'Regression Learner App' of MATLAB. The training data was comprised of four predictor variables i.e. 'BPFI Amplitude', 'BPFO Amplitude', log ratio {i.e.log $\left(\frac{BPFIAmplitude}{BPFOAmplitude}\right)$} and bearing fault stages (class labels); the response variable being the average RUL corresponding to each of the fault stages (class labels). The performance and accuracy of the developed liner regression ML model was mainly dependant on the allocation of class labels by the classification ML model i.e. Weighted KNN as the average RUL values (in hours) were trained corresponding to the allocated fault stage. Figure 9a shows the response plot of linear regression ML model trained with the same training data as that used to develop Weighted KNN model. Figure 9b shows the response plot of the same regression trained model with respect to log ratios.

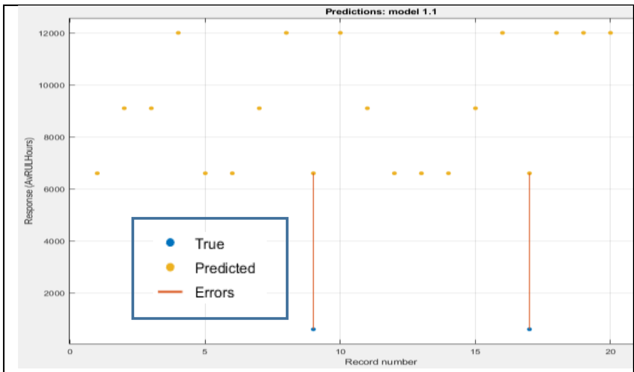


Figure 9a: Response plot of linear regression ML model trained to estimate average RUL corresponding to the classifications allocated by Weighted KNN model

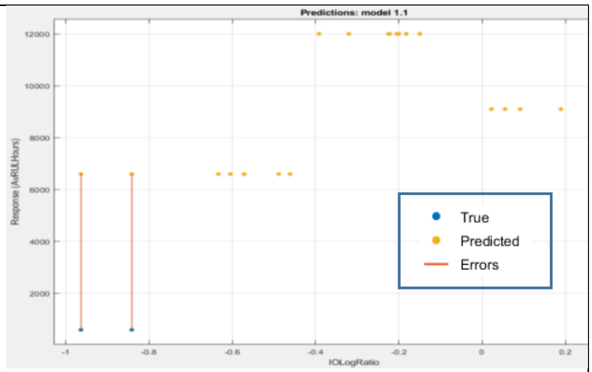


Figure 9b: Response plot (IOLogRatio vs AvRULHours) of linear regression ML model trained with the same training dataset as that used to train Weighted KNN model

Figure 9: Response plots of linear regression ML model

5 Conclusion

Similar to its application for many other data driven systems, processes and operations, ML can also be a game changer for the performance optimisation of an Asset Management System and thus enhancing the value of asset management tangibly or intangibly. The approach of transforming human expert knowledge into ML models can be applied effectively to capture and preserve the knowledge, experience and expertise of humans for any process and/or analysis. Such a ML model with high level of accuracy can be utilised not only to improve the process control due to a tried-and-tested decision making framework but number of humans required to do the job can also be reduced significantly. Furthermore, such ML models also provide an opportunity to improve the training standards of future humans and to define new benchmarks of system and equipment design.

Due to a number of specific challenges involved pertaining to the logistics of data acquisition for a defence maritime materiel, a ML model inculcated with human expert knowledge can play a pivotal role in developing a decision making framework for the adoption and implementation of risk based maintenance philosophies such as PdM and CBM etc. For instance, such ML models can employed effectively to make more accurate and timely decisions on the replacement of parts in order to minimise the failures at sea and reducing their replacement earlier that required, thus enhancing the availability and affordability of the materiel respectively.

ML models can be developed for both diagnostic and prognostic analysis of the systems, equipment and their components. However, similar to the diversity of techniques used in data processing, there is no pre-determined best method, algorithm or technique which can be applied to any specific real world scenario even, if a similar scenario has previously been analysed. Every application of ML is unique and, therefore, significant investment of time and effort is required by appropriately skilled people to achieve the best possible application of a ML model.

The more a model is trained with a variety of datasets, the more it will generalise the given process and, thus, the more meaningfully it will produce the results.

The next version of the developed ML model can include the load of the rotating machines as one of the predictor variables. The application of ML is not limited to the condition monitoring only, but it can be very effectively applied for the operations of the machinery wherein the critical operational parameters of the machinery specific to the system environment can be used as CIs to establish the diagnostic and prognostic analysis.

Acknowledgements

The Authors would like to extend my special thanks to ASC Pty Ltd and its management for sponsoring this research project and providing me with the right resources, training and real time data to make this research more meaningful and effective.

6 References

A Comparative Study on Bearings Faults Classification by Artificial Neural Networks and Self-Organizing Maps using Wavelets. **Kalyan M. Bhavaraju, P. K. Kankar, Satish C. Sharma, S. P. Harsha. 2010.** 5, Roorkee : MultiCraft, 2010, International Journal of Engineering Science and Technology, Vol. 2, pp. 1001-1008. ISSN: 0975-5462.

An Overview on Vibration Analysis Techniques for the Diagnosis of Rolling Element Bearing Faults. **Patidar, Shyam and Soni, Pardeep. 2013.** 5, Bhopal : s.n., May, 2013, International Journal of Engineering Trends and Technology (IJETT), Vol. 4. ISSN: 2231-5381.

Ball Bearing Fault Diagnosis Using Supervised and Unsupervised Machine Learning Methods. **V. Vakharia, V. K. Gupta and P. K. Kankar. 2015.** 4, Jabalpur : International Institute of Acoustics and Vibration (IIAV), 2015, The International Journal of Acoustics and Vibration (IJAV), Vol. 20, pp. 244-250. ISSN: 2415-1408.

Bearing Envelope Analysis Window Selection. **Bechhoefer, Dr. Eric and Menon, Praneet. 2009.** Vergennes : Goodrich – Sensors & Integrated Systems, 2009.

Bearing Time-to-Failure Estimation using Spectral Analysis Features. **Keong, Reuben Lim Chi and Mba, David. 2014.** 2, Cranfield : SAGE Publishing, 2014, Structural Health Monitoring, Vol. 13, pp. 219-230. eISSN: 17413168.

Brown, Peter. *Fundamentals of Vibration Measurement and Analysis Explained.* [Document] Perth : Lifetime Reliability Solutions.

CrowdFlower, Inc. 2016. *206 Data Science Report.* San Francisco : CrowdFlower, Inc., 2016. p. 10, Survey Report.

Data mining and machine learning for condition-based maintenance. **Accorsi, Riccardo, et al. 2017.** 2017, Procedia Manufacturing, Vol. 11, pp. 1153 – 1161.

Fault diagnosis of rolling element bearings via discriminative subspace learning: Visualization and classification. **Mingbo Zhao, Xiaohang Jin, Zhao Zhang, Bing Li. 2014.** 7, Suzhou : Elsevier, 2014, Expert Systems with Applications, Vol. 41, pp. 3391-3401.

Global Forum on Maintenance and Asset Managment (GFMAM). 2014. The Asset Managment Landscape - Second Edition. Zurich : Global Forum on Maintenance and Asset Managment (GFMAM), 2014.

Han Jiawei, Kamber Micheline, Pei Jian. 2012. Chapter 2 - Getting to Know Your Data. *Data Mining - Concepts and Techniques.* 3rd Ed. Waltham : Morgan Kaufmann Publishers-ELSEVIER, 2012, 1, pp. 39-81.

—. 2012. Chapter 3 - Data Preprocessing. *Data Mining - Concepts and Techniques.* 3rd Ed. Waltham : Morgan Kaufmann Publishers-ELSEVIER, 2012, 1, pp. 39-81.

Howieson, Donald. 2003. *Vibration Monitoring:Envelope Signal Processing.* [Document] San Diego : SKF Reliability Systems, 2003.

Liu, Fenfen. 2015. *Condition Monitoring and Prognosis for Subsea Multiphase Pump.* Department of Production and Quality Engineering, Norwegian University of Science and Technology. Trondheim : Norwegian University of Science and Technology, 2015.

Mais, Jason. 2002. *Spectrum Analysis - The key features of analyzing spectra.* [Document] San Diego : SKF USA Inc., 2002. CM5118 EN.

Marouani, Stéphane. 2018. Can engineers help fill the data scientist gap? *Create - Engineering Ideas into Reality.* [Online] Engineers Australia, 14 Nov, 2018. [Cited: 30 Nov, 2018.] https://www.createdigital.org.au/engineers-fill-data-scientist-gap/.

MathWorks Inc. 1994-2018. Design and test condition monitoring and predictive maintenance algorithms. *MathWork.* [Online] MathWorks Inc., 1994-2018. [Cited: 03 October, 2018.] Web Page. https://au.mathworks.com/products/predictive-maintenance.html?s_tid=srchtitle.

—. 1994-2018. Overcoming Four Common Obstacles to Predictive Maintenance with MATLAB and Simulink. *MathWorks.* [Online] 1994-2018. [Cited: 3 October, 2018.] https://au.mathworks.com/content/dam/mathworks/white-paper/gated/predictive-maintenance-challenges-whitepaper.pdf.

—. 1994-2018. Three Ways to Estimate Remaining Useful Life for Predictive Maintenance. *MathWorks.* [Online] 1994-2018. [Cited: 15 October, 2018.] https://au.mathworks.com/content/dam/mathworks/ebook/estimating-remaining-useful-life-ebook.pdf.

MathWorks, Inc. 1994-2019. Assess Classifier Performance in Classification Learner. *MathWorks, Inc.* [Online] MathWorks, Inc., 1994-2019. [Cited: 23 January, 2019.] https://au.mathworks.com/help/stats/assess-classifier-performance.html.

—. 1994-2019. Choose Classifier Options. *MathWorks, Inc.* [Online] MathWorks, Inc., 1994-2019. [Cited: 2 January, 2019.] https://au.mathworks.com/help/stats/choose-a-classifier.html.

MathWorks, Inc. 1994-2018. Data Ensembles for Condition Monitoring and Predictive Maintenance. *MathWorks.* [Online] MathWorks, Inc., 1994-2018. [Cited: 5 October, 2018.] https://au.mathworks.com/help/predmaint/ug/data-ensembles-for-condition-monitoring-and-predictive-maintenance.html.

MathWorks, Inc. 1994-2019. Designing Algorithms for Condition Monitoring and Predictive Maintenance. *MathWorks.* [Online] MathWorks, Inc., 1994-2019. [Cited: 05 January, 2019.] https://au.mathworks.com/help/predmaint/gs/designing-algorithms-for-condition-monitoring-and-predictive-maintenance.html.

—. 2016. *Getting Started with Machine Learning.* [Document] Natick, Massachusetts, United States : MathWorks, Inc., 2016.

—. 2016. *Introducing Machine Learning.* [Document] Natick, Massachusetts, United States : MathWorks, Inc., 2016.

—. 2018. Machine Learning with MATLAB - Training Course. *Machine Learning with MATLAB.* [Training Manual]. Natick, Massachusetts, United States : MathWorks Training Services, 2018. Vol. 1, MLML418.

—. 2018. *Mastering Machine Learning - A Step-by-Step Guide with MATLAB.* [e-book] Natick, Massachusetts, United States : MathWorks, Inc., 2018.

—. 1994-2019. Predict Remaining Useful Life. *MathWorks.* [Online] MathWorks, Inc., 1994-2019. [Cited: 5 January, 2019.] https://au.mathworks.com/help/predmaint/predict-remaining-useful-life.html.

—. 1994-2018. Rolling Element Bearing Fault Diagnosis. *MathWorks.* [Online] MathWorks, Inc., 1994-2018. [Cited: 23 April, 2018.] https://au.mathworks.com/help/predmaint/ug/Rolling-Element-Bearing-Fault-Diagnosis.html.

—. 1994-2018. Train Classification Models in Classification Learner App. *MathWorks.* [Online] MathWorks, Inc., 1994-2018. [Cited: Sep 12, 2018.] https://au.mathworks.com/help/stats/train-classification-models-in-classification-learner-app.html.

MatWorks, Inc. 1994-2019. RUL Estimation Using RUL Estimator Models. *MathWorks.* [Online] MathWorks, Inc., 1994-2019. [Cited: 18 December, 2018.] https://au.mathworks.com/help/predmaint/ug/rul-estimation-using-rul-estimator-models.html.

Mobius Institute. 2012. *Vibration Training Quick Reference.* Melbourne : Mobius Institute, 2012.

Moubray, John. 1997. *Reliability Centered Maintenance-II.* Second. Massachusetts : Butterworth-Heinemann, 1997.

National Institute of Standards and Technology (NIST) . 2013. Introduction to Time Series Analysis. *Engineering Statistics Handbook.* [Online] US Department of Commerce , 10 October, 2013. [Cited: 22 Augusat, 2018.] https://www.itl.nist.gov/div898/handbook/index.htm.

Oracle Corporation. 2018. What Is Big Data? *Oracle Australia.* [Online] Oracle Corporation, 2018. [Cited: 9 January, 2019.] https://www.oracle.com/au/big-data/guide/what-is-big-data.html.

Press, Gil. 2015. 6 Predictions For Big Data Analytics And Cognitive Computing In 2016. *Forbes.* [Online] Forbes, 15 Dec, 2015. [Cited: 12 Dec, 2018.] https://www.forbes.com/sites/fastforward/2019/01/08/these-founders-killed-it-in-2018-and-their-returns-are-all-impact/#7b6448ae77a1.

ReliaSoft Corporation. 2015. *System Analysis Reference.* [Document] Tucson, Arizona, USA : ReliaSoft Corporation, 5 May, 2015.

Remaining Useful Life Prediction of Rolling Bearings Using PSR, JADE, and Extreme Learning Machine. **Yongbin Liu, Bing He, Fang Liu, Siliang Lu, Yilei Zhao, Jiwen Zhao. 2016.** Hefei : Hindawi Publishing Corporation, 2016, Mathematical Problems in Engineering, Vol. 2016.

Risk-Based Maintenance Scheduling with application to naval vessels and ships. **Cullum, Jane, et al. 2018.** 2018, s.l. : Elsevier, 2018, Ocean Engineering, Vol. 148, pp. 476–485. ISSN: 0029-8018.

Rolling Element Bearing Analysis. **Graney, Brian and Starry, Ken. 2012.** 1, Ohio : The American Society for Nondestructive Testing, Inc., 2012, Materials Evaluation, Vol. 70, pp. 78-85.

Rolling element bearing diagnostics—a tutorial. **Randall, B, Robert and Antoni, Jerome. 2011.** 2, Amsterdam : Elsevier, 2011, Mechanical Systems and Signal Processing, Vol. 25, pp. 485-520.

Standards Australia Limited. 2014. AS ISO 55000:2014 - Asset management—Overview, principles and terminology. Sydney : Standards Australia Limited, 2014.

Vachtsevanos, G, et al. 2006. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems.* New Jersey : John Wiley & Sons, Inc, 2006.

Willingham, David. 2015. *Machine Learning Made Easy.* [Presentation] Natick, Massachusetts, United States : MathWorks, Inc., 2015.
Witten, Ian H., Frank, Eibe. 2005. *Data Mining - Practical Machine Learning Tools and Techniques.* 2nd. San Francisco : Morgan Kaufmann Publishers, 2005. ISBN: 0-12-088407-0.
Xu, Ming. 1998. *Spike Energy Measurement and Case Histories.* [Paper] Ohio : Rockwell Automation, Integrated Condition Monitoring, Westerville, Ohio, 1998.
Zhou, Jianlong and Chen, Fang. 2018. Chapter 1 -2D Transparency Space - Bringing Domain Users and Machine Learning Experts Together. [book auth.] Behnoush Abdollahi, et al. *Human and Machine Learning.* Cham : Springer International Publishing AG, 2018, pp. 3-19.