MINITRY OF EDUCATION AND TRAINING

HCMC UNIVERSITY OF TECHNOLOGY

AND EDUCATION

FACULTY OF MECHANICAL ENGINEERING

HCMUTE

REPORT

DIGITAL SIGNAL PROCESSING

REVIEW TECHNOLOGIES USED TO ANALYZE VIBRATION SIGNALS FROM BEARINGS AND GEARBOXES

Lecturer: Bui Ha Duc

Students report:

	Full Name	ID
1	Tran Nhat Hoang	21134008
2	Lai The Trung	21151405

Ho Chi Minh City, December 2022

End-Semester Report

ABSTRACT

Nowadays we are surrounded by machines. Most of our tasks are get completed by rotatory machines. And faults in machines are natural things. Rotating mechanical components in machinery like bearings, gears, pulleys, belt drives, etc. are major components in any rotating machinery. The failure of these components leads to the downtime of machines and reduction in production. Significant economic losses will be caused due to an unexpected failure of these components. Belt drives are widely employed in various industrial equipment. Finding the early fault symptoms in the belt drive is very important. This can be achieved by various methods. For detecting faults and monitoring the condition of a belt drive, the vibration signal can be used as one of the parameters. Thus, vibration signal can be used as a procedure for predictive maintenance and it is used for machinery maintenance decisions. The changes in vibration signals due to fault can be detected by employing signal processing methods. It can be used to evaluate the health status of the machinery. The nature and severity of the problem can be determined by analyzing the vibration signal and hence the failure can be predicted. The signature of the fault in the machine is carried by the vibration signal. It is possible to have early fault detection by analyzing these vibration signals. Different signal processing techniques are used for processing these signals. The various techniques used for fault diagnosis based on the vibration analysis method are discussed in this paper. The application of artificial intelligence techniques such as Artificial Neural Networks (ANN), fuzzy sets, and other emerging technologies are discussed.

I. INTRODUCTION

1. Aim of the project:

The aim of this project is to explain the functioning and the efficiency of the techniques of vibration analysis used in the predictive plan.
In this project, we just focus on the techiques and method, the suitable ways to analysis vibration signal from industrial and then put all together to machine learning network to learn if the gearbox have fault or not.

2. The predictive maintenance techniques and why vibration analysis:

Nowadays, there is a vast maintenance techniques spectrum to choose. In a proper maintenance plan at least two techniques are applied, as the use of only one sometimes does not provide a clear equipment diagnosis. This decision should be made taking into account the type of industry, the equipment used in the manufacturing process, the competent personnel available, and the formation the staff needs to carry out these techniques.

The most common ones are listed and briefly explained below:

- Acoustic emission: this technique can be useful to locate cracks in structures and pipelines.
- Lubrication oil and wear particles analysis: the equipment condition is controlled thanks to a study in the oil composition.
- Infrared thermography: very efficient at detecting failure in electric and insulating elements, as it measures the temperature of the studied element.
- Ultrasound scanning: this technique is employed to detect leakages and to check corrosive wear levels.
- Vibration analysis: it has been demonstrated that, among all other non-destructive techniques, this one provides the greatest amount of information among all others.

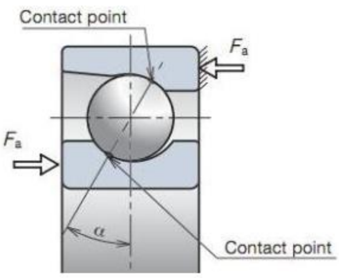
However, As it has been said that the most complete analysis of the studied element condition, is obtained through a vibration analysis. This technique is proved to be the most effective way to detect rotating elements failure. As a consequence of the wear of the machine's elements, vibrations begin to propagate all along the equipment, which leads to additional forces application. This vibration is measured in those points that put up with them.

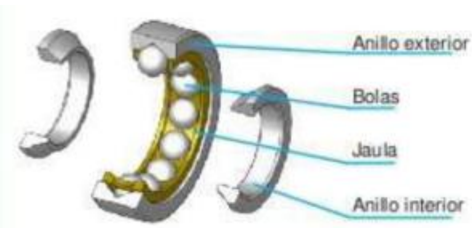
One of the most remarkable advantages of this method is the ability of identifying failure symptoms before this failure appears. This technique can be employed to detect deterioration in bearings and gears, as well as misalignment and unbalance symptoms, before these result in a breakdown.

When using vibration analysis techniques, the state of the machinery is determined by studying the characteristics of its vibration, being the most important ones the frequency and the amplitude.

Depend on the extraodinary points in the amplitude or frequency graph, we can access the problems in rotating machine.

Example for the powerful of frequency and amplitude on vibration signals analysis, the graph of amplitude- frequency on the left side corresponding to the image of fault systems on the right side.





For each problem to the bearing ring, we receive different signal of vibration

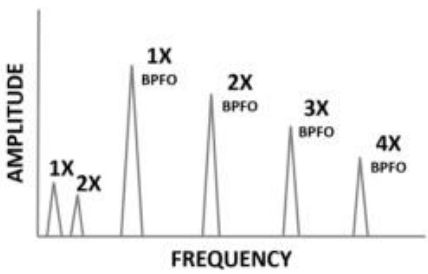


Fig. 24: Spectrum corresponding to a defect in the outer ring

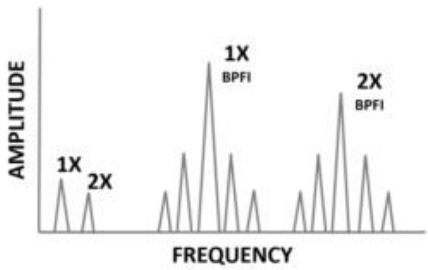
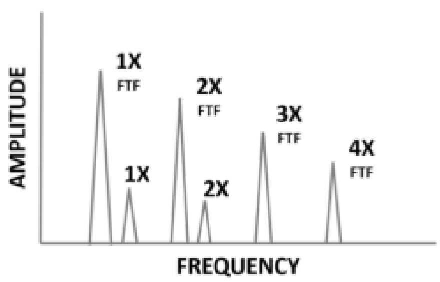


Fig. 25: Spectrum corresponding to a defect in the inner ring



F

Fig. 27: Spectrum corresponding to a defect in the cage

II. State of Art methods for 1D Vibration data analysis

1. Time domain method:

a) Ideal:

Time domain analysis usually involves scalar indices to determine the bearing condition through temporal vibrational signal data and based on its value bearing condition can be estimated.

b) Feature:

Some approach for feature extraction in time domain is :

Approach 1: Apply on raw data:

Feature Name	Description	
	Brief Definition	Formula
RMS	The RMS value increase gradually as fault developed. However, RMS is unable to provide the information of incipient fault stage while it increases with the fault development [11].	$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Variance	Variance measures the dispersion of a signal around their reference mean value.	$\mathrm{Var} = \frac{\sum_{i=1}^{N}(x_i - m)^2}{(N-1)\sigma^2}$
Skewness	Skewness quantifies the asymmetry behavior of vibration signal through its probability density function (PDF).	$\mathrm{Sk} = \frac{\sum_{i=1}^{N}(x_i - m)^3}{(N-1)\sigma^3}$
Kurtosis	Kurtosis quantifies the peak value of the PDF. The kurtosis value for normal rolling element bearing is well-recognized as 3.	$\mathrm{Ku} = \frac{\sum_{i=1}^{N}(x_i - m)^4}{(N-1)\sigma^4}$
Shape factor	Shape factor is a value that is affected by an object's shape but is independent of its dimensions [12].	$\mathrm{SF} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}}{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert}$
Crest factor	Crest factor (CF) calculates how much impact occur during the rolling element and raceway contact. CF is appropriate for "spiky signals" [12].	$\mathrm{CF} = \frac{\max\lvert x_i\rvert}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}}$
Entropy	Entropy, $e(p)$, is a calculation of the uncertainty and randomness of a sampled vibration data. Given a set of probabilities, $(p_1, p_2, \ldots, p_n)$, the entropy can be calculated using the formulas as shown in the right column.	$e(p) = -\sum_{i=1}^{n} p(z_i)\log_2 p(z_i)$

RMS : The simplest approach in the time domain is to measure the overall root-mean-square level and crest factors to identify the differences between one vibration signal and another. This method has been applied with limited success for the detection of localized defects (includes cracks, spalls in the rolling surface)

Skewness and kurtosis can be applied to the signal which is not purely stationary. These features examine the probability density function (PDF) of the signal.Kurtosis measures the peak value of the PDF and indicates if the signal is impulse in nature while skewness is used to measure whether the signal is negatively or positively skewed "Skewness coefficient " used to normalized with respect to the cube of standard deviation. These features examine the probability density function (PDF) of the signal.

Entropy, which calculates the histogram of the PDF and measures the degree of randomness of the vibration signal.

Crest factor is simply the ratio of the peak acceleration to the RMS acceleration, so it is unitless which is always nice. It defines how "peaky" a signal is.

Lower bound and upper bound histogram, which measure the lower and upper values of the PDF respectively

In Figure 1, some features, such as RMS, kurtosis, histogram upper bound and histogram lower bound, show the fluctuation in the last measurement day, approximately from the 90th day to the 139th day. These indicate that some features are sensitive to the slew bearing condition while others are less.

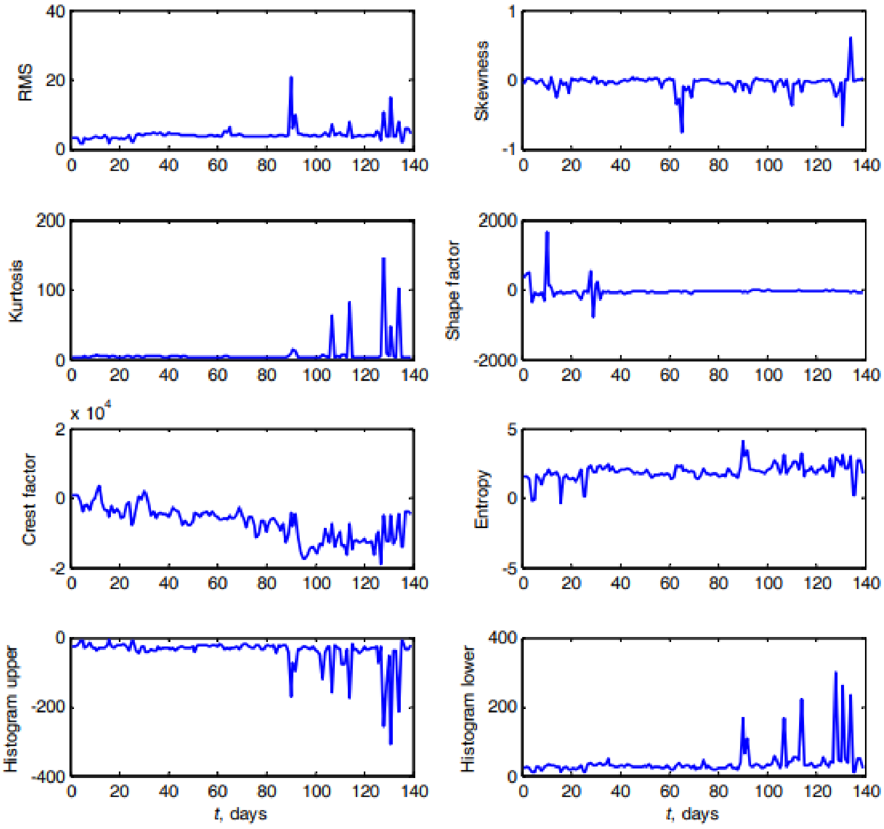


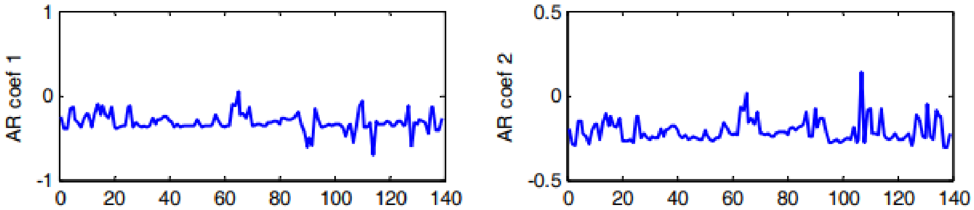
Figure1: Statistical time-domain features calculated from the vibration slew bearing signal.

These features only shows the damage at the ball bearing but do not give information about the location of defect e.g. inner race, outer race, cage or the roller.

<u>Approach 2: Apply extraction on data with special coefficient:</u>

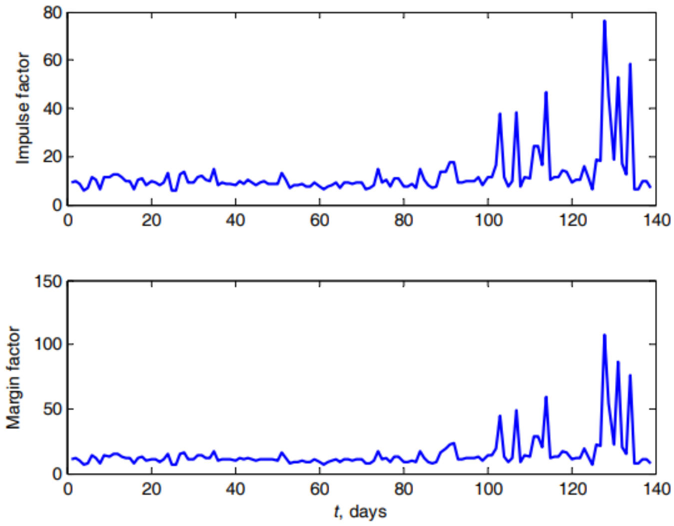
Autoregressive (AR) Coefficients: of the vibration signal are calculated as bearing vibration features. It is known that a faulty vibration signal of typical rolling element bearing produces different autoregressive coefficients compared to that of normal vibration signal.

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_n y_{t-n} + \varepsilon_t = \sum_{i=1}^{n} a_i y_{t-1} + \varepsilon_t$$



It is shown that most of the features are less relevant to represent the slew bearing condition except for AR coefficient 4 which is shown fluctuating in the last measurement.

Other time domain features such as impulse factor (IF) and margin factor (MF) have been recently used. IF measure how much impact is generated from the bearing defect while MF measures the level of impact between rolling element and raceway.



The progression of slew bearing deterioration of IF and MF features are more apparent compared to RMS, kurtosis and histogram lower features.

Hjorts' Parameters :In addition to the time-domain features mentioned in previous subsections, Hjorts' parameters also falls into this category. These features are calculated based on the first and the second derivatives of the vibration signal. In the time series context, the numerical values for the derivatives are obtained as the differences between the current value and the prior value.

These includes 2 parameters (where σx' is the standard deviation of the first derivative of the vibration signal)

$$\text{mobility} = \frac{\sigma_{x'}}{\sigma_x} \qquad \text{complexity} = \frac{\sigma_{x''}/\sigma_{x'}}{\sigma_{x'}/\sigma_x}$$

However, They have never been used in vibration bearing signal except for activity feature, which is similar to the variance feature in the statistical time-domain features extraction.The progression of bearing condition is difficult to track using mobility and complexity.

Mathematical Morphology (MM) Operators: The fundamental principle of the MM method is to modify the shape of the original signal by transforming it through its intersection with another object called the 'structuring element' (SE). Four operators including erosion, dilation, closing and opening are used to achieve the transformation.

- *Erosion*: also refer to as min filter.

$$(x \Theta S)(n) = \min(x(n+m) - S(m)) \tag{11}$$

- *Dilation*: also refer to as max filter.

$$(x \oplus S)(n) = \max(x(n-m) + S(m)) \tag{12}$$

- *Closing*: Dilates 1D signal and then erodes the dilated signal using the similar structuring element for both operations.

$$(x \bullet S)(n) = ((x \oplus S)\Theta S)(n) \tag{13}$$

- *Opening*: Erodes 1D signal and then dilates the eroded signal using the similar structuring element for both operations.

$$(x \circ S)(n) = ((x \Theta S) \oplus S)(n) \tag{14}$$

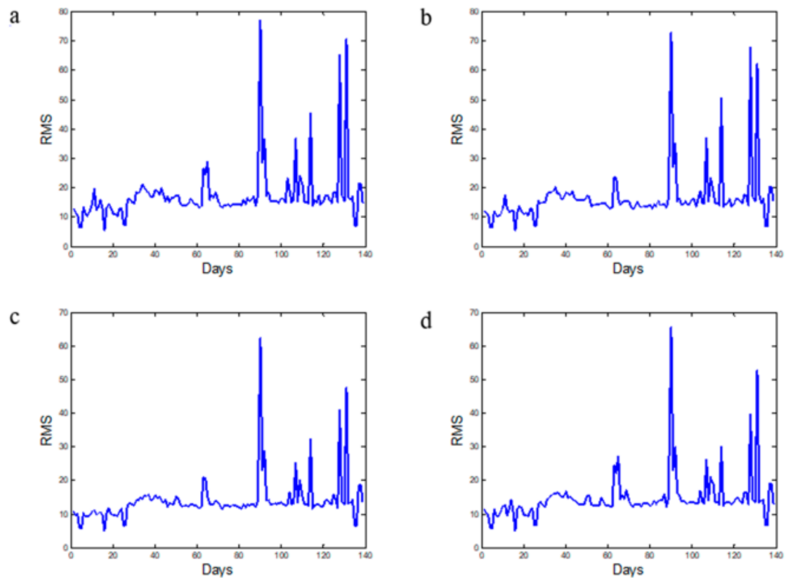


Figure 2: RMS feature calculated from MM operators for 0.9T: a) erosion; b) dialation; c) closing; d) opening

The MM method has been used effectively for the damage detection of high speed rollingbearing. It is because when the damage occurred, the contact between the damage spot and rollingelement generates detectable impulse due to the contact between the defect spot and race way orroller. In the case ofnaturally degraded low-speed slew bearing, the multiple damage e.g., outer race,Inner race and roller are typically occurred at close sequence time and the magnitude of impulse is very low. Thus, it is difficult to identify the origin of impulse from outer race, inner race.

c) Advantage and disadvantage:

Time domain analysis is the simplest technique for fault detection in bearings. Detect well in stationary, high speed rotation, high frequency.

However, It only shows the damage at the ball bearing but do not give information about the location of defect and seems not too much effective in low speed and frequency detection. when the damage progresses vibrational signals become more and more random, the values of these features seems hard for distinguishing.

2. Frequency Domain method:

a) Ideal:

Frequency domain or spectral analysis is the most widely used approach for fault diagnosis in bearings. Frequency-domain techniques convert time-domain vibration signals into discrete frequency components using a fast Fourier transform (FFT).

b) <u>Feature:</u>

Fault signal can be identified based on bearing fundamental frequencies which depends on the kinematic consideration i.e. bearing geometry and rotor speed. For a bearing with stationary outer race, these frequencies have been calculated using the following expressions

$$\textit{Inner Race Ball Pass Frequency (BPFI)} = \frac{n}{2} f_r \left(1 + \frac{d_b}{d_p \mathrm{Cos}\varphi}\right) \tag{2}$$

$$\textit{Outer Race Ball Pass Frequency (BPFO)} = \frac{n}{2} f_r \left(1 - \frac{d_b}{d_p \mathrm{Cos}\varphi}\right) \tag{3}$$

$$\textit{Ball Pass Spin Frequency (BSF)} = \frac{n}{2} f_r \left(1 - \frac{d_b^2}{d_p^2 \mathrm{Cos}^2\varphi}\right) \tag{4}$$

$$\textit{Fundamental train frequency (FTF)} = f_r \left(1 - \frac{d_b}{d_p \mathrm{Cos}\varphi}\right) \tag{5}$$

Where n = number of balls in rolling element bearing, d_b = ball diameter, d_p = bearing pitch diameter, φ = bearing contact angle, f_r = rotor frequency.

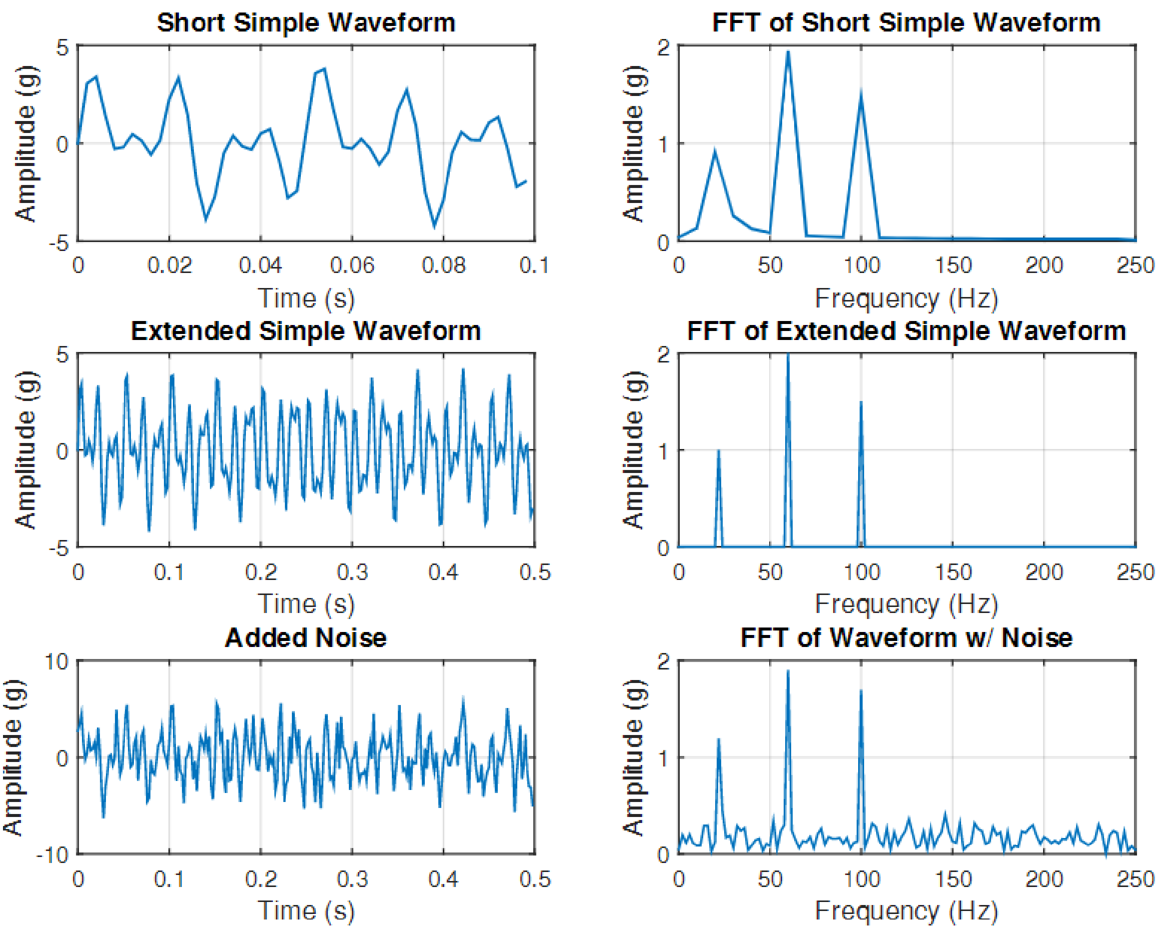
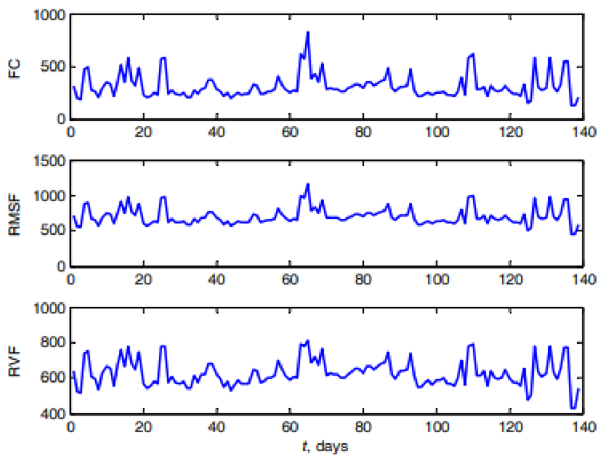


Figure 3: The frequency domain from the data of a vibration sensor, clearly seen the composite frequency of the raw wave signal.

Analysis of raw vibration signal in frequency domain can be done through Discrete Fourier transform (DFT) and Fast Fourier transform (FFT). DFT is a straight mathematical procedure but is inefficient while fast Fourier transform (FFT) analysers are an easier and efficient way of obtaining narrow band spectra. In a frequency spectrum plot, the X-axis represents frequency and the Y-axis represents amplitude of displacement, velocity, or acceleration.

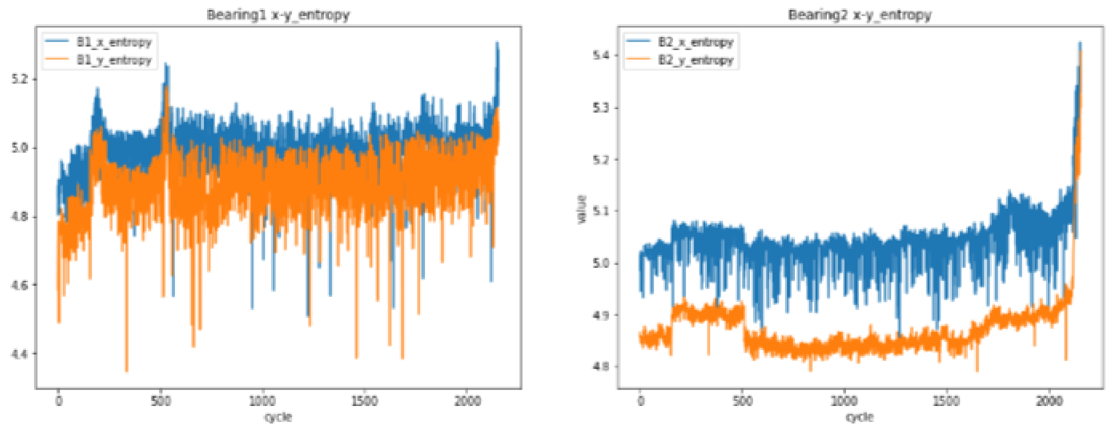
From the data take from Frequency domain, applied simple extraction frequency centre (FC), root mean square frequency(RMSF) and root variance frequency (RVF) to get the Statistical Frequency-Domain Features or analysis.

$$\mathrm{FC} = \frac{\sum_{i=2}^{N} x_i' x_i}{2\pi \sum_{i=1}^{N} x_i^2} \qquad \mathrm{MSF} = \frac{\sum_{i=2}^{N} (x_i')^2}{4\pi^2 \sum_{i=1}^{N} x_i^2} \qquad \mathrm{RMSF} = \sqrt{\mathrm{MSF}}$$



Recently, new methods such as spectral skewness, spectral kurtosis, spectral entropy and Shannon entropy have been developed. Spectral skewness (SS) and spectral kurtosis (SK) are the advanced statistical measures applied to the magnitude spectrum.

$$\mathrm{SS}(n) = \frac{2\sum_{k=0}^{\mathrm{B}_L/2-1}\left(\left|X(k,n)\right| - \mu_{|X|}\right)^3}{\mathrm{B}_L \cdot \sigma_{|X|}^3} \qquad \mathrm{SK}(n) = \frac{2\sum_{k=0}^{\mathrm{B}_L/2-1}\left(\left|X(k,n)\right| - \mu_{|X|}\right)^4}{\mathrm{B}_L \cdot \sigma_{|X|}^4} - 3$$



The concept of spectral statistic is adopted in this method to supplement the classical power spectral density (PSD). PSD ideally gives zeros values at those frequencies when the signal has stationary Gaussian noise, and it gives high positive values at those frequencies during the occurrence of the transients. When the noise-to-signal ratio is high, the transient signal will be buried in the background noise, thus the incipient fault cannot be clearly detected. SK can overcome this problem by analyzing the entire frequency band and to select the sensitive frequency band that correspond to the bearing condition.

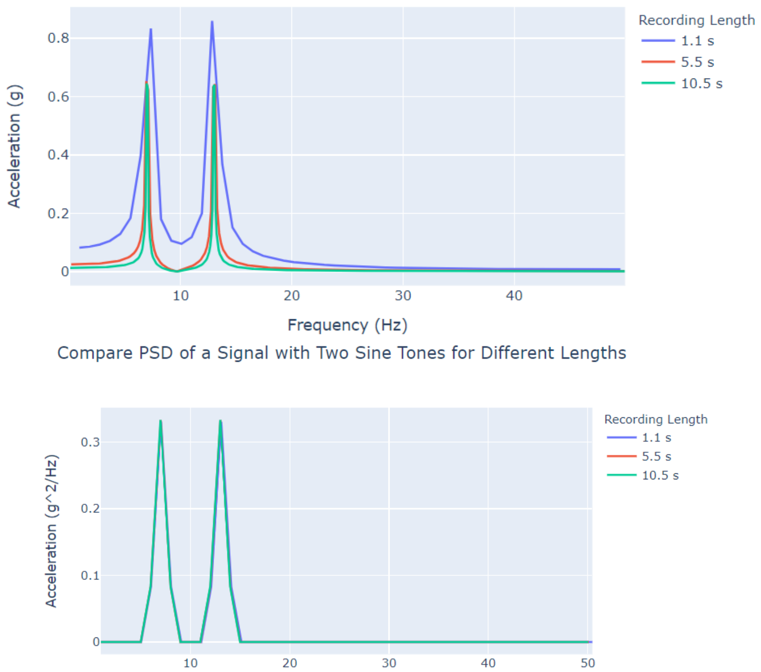


Figure 4: the PSD of different signal lengths just fills in this area but the amplitude doesn't change overall. The FFT amplitude however shifts down as the bandwidth is increased.

c) Advantage and disadvantage:

Frequency domain technique is capable of detecting certain frequency components of interest very easily which is an advantage over time domain technique. Well described the order of appearing and disappearing of peaks in the spectrum. Advantage over time domain is it can identify defects as well as the defect location.

FFT is effective when signals are non-stationary and signal to noise ratio is low. But with high noise signal, FFT works not well.

3. Time-Frequency Domain method:

a) Ideal:

In rotating machines, bearing vibration signal is a combination of periodic components, dominated by the machine rotation, with signals of a random nature, dominated by a possible bearing fault or imperfection. This phenomenon is periodic in nature, has time invariance and non-stationary. This has led to the development of time-frequency analysis methods

b) Feature:

Short-Time Fourier Transform (STFT): STFT divides a non-stationary signal into small windows of equal time frame. Thewindow is applied to each segment while also overlapping all segments so that we aren't filtering anything away .The Fourier transformation is then applied to the time segment.

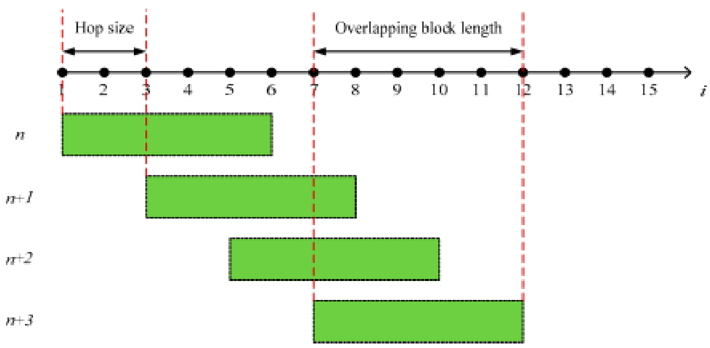


Figure 5: Ideal of STFT visualized

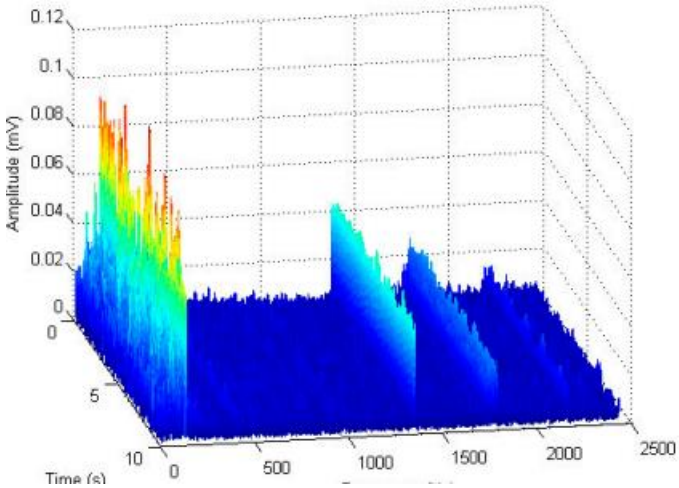


Figure 6: From 1D data, SFFT give 2D data.

Applied Feature staticextraction for these 2D dataset to get Statistical Frequency-Domain Features and Time-Domain Features.

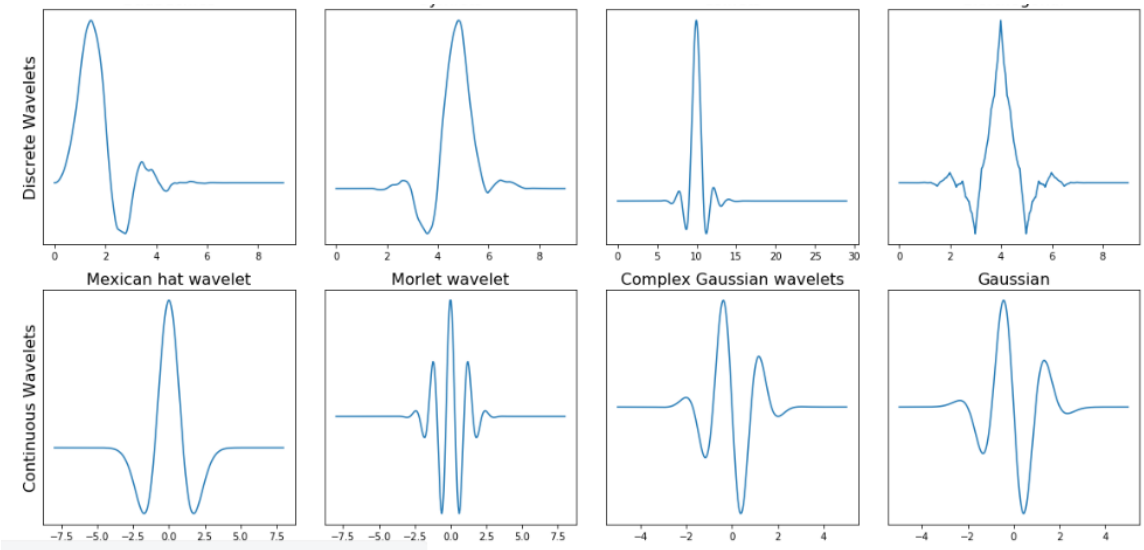
Wigner-Ville Distribution (WVD): Wigner-Ville Distribution is a Fourier transform of a product of the signal with its complex conjugate. Thus, it resembles the computation of the power spectral density. The WVD has been used for gear fault detection [47,48] and it is recently used for rolling element bearing to represent the time-frequency features of vibration signals.

Wavelet Transform and Wavelet Decomposition: a is the most popular and widely used time–frequency domain technique for bearing defect detection. A waveform of limited duration that has an average value of zero. In wavelet analysis, signal is broken up into shifted and scaled versions of the original wavelet. For the diagnosis of fault in machines, the main features of wavelets are time–frequency analysis of signals, fault feature extraction, detection of singularity for signals, extracting weak signals, compression of vibration signals and the system identification . Because wavelet is flexible and implemented effectively in computation, it is a very effective tool for machine condition monitoring and fault diagnosis and bearing fault detection in particular. The wavelet function is defined as:

$$\mathrm{WT}_x(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(\tau)\psi\left(\frac{\tau - t}{a}\right) d\tau$$

Wavelet transform are continuous and discrete wavelet type. Continuous Wavelet Transform is product of signal with translate and dilate of wavelet. CWT uses short windows at high frequencies and long windows at low frequencies while

Discrete Wavelet Transform (DWT) a small number of scales are used to analyse signal and number oftranslation vary at each scale.



The wavelet transform can represent the signal with a certain number of coefficients. The process is usually called wavelet decomposition.

The structure illustrates the decomposition of slew bearing vibration signal into approximate coefficient (A) and detailed coefficient (D) at each level. For further processing, features such as mean, variance, skewness and kurtosis are calculated from the detailed coefficient at each level.

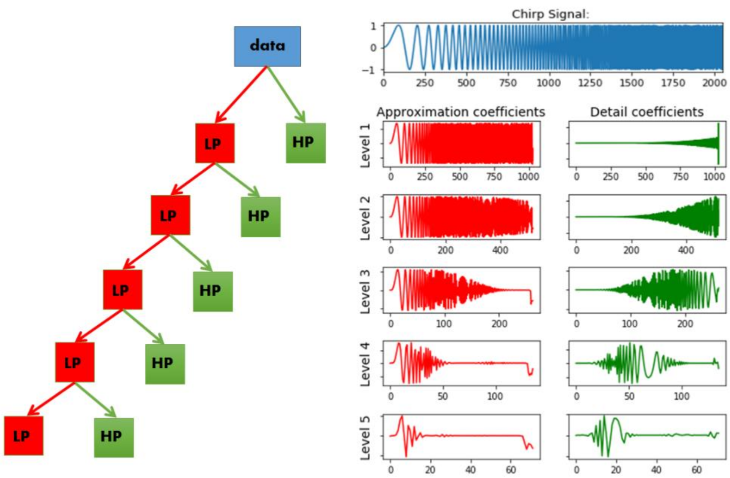


Figure 7: Mapping Wavelet Decomposition

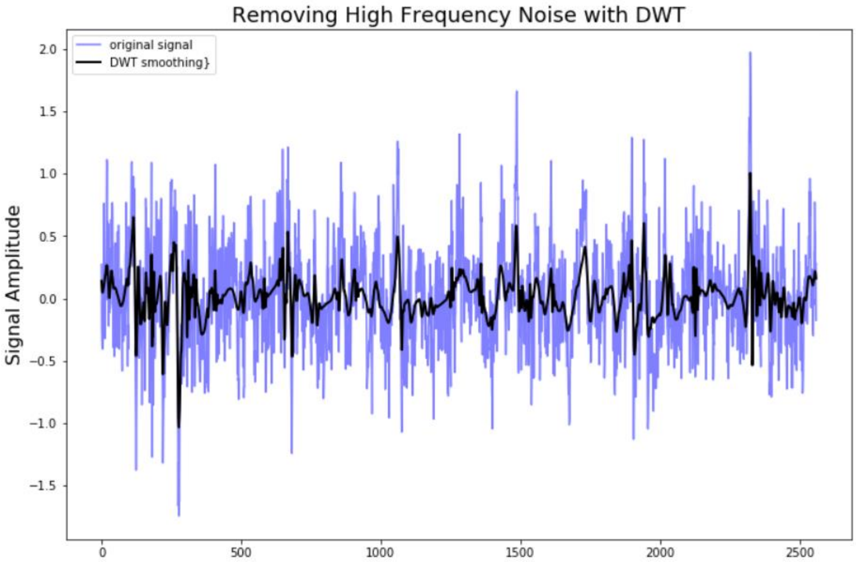


Figure 7: Example for applying Discrete Wavelet Decompostion to denoise

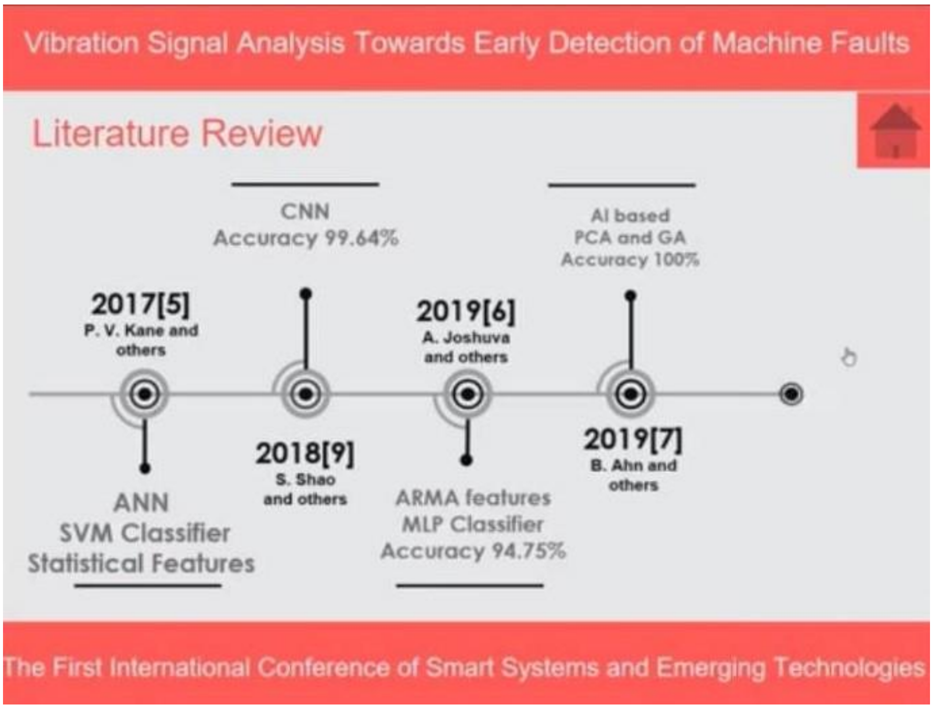
Applied Feature staticextraction for these 2D dataset to get Statistical Frequency-Domain Features and Time-Domain Features

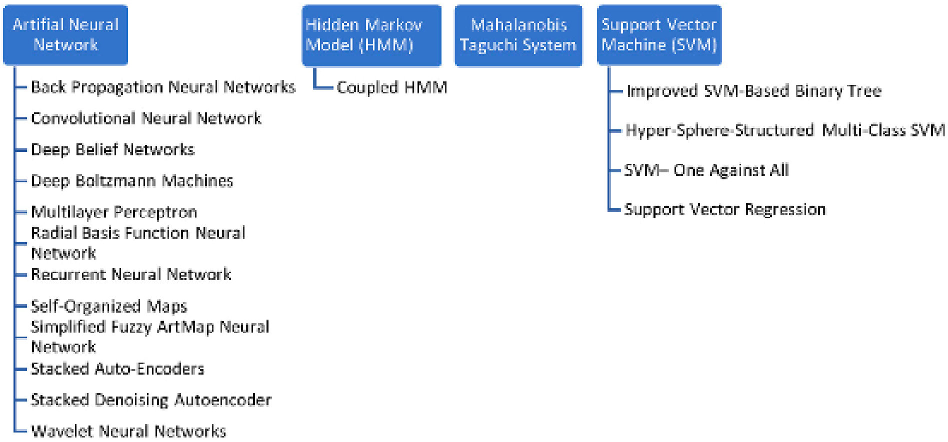
c) <u>Advantage and disadvantage:</u>

The advantage of time-frequency domain techniques over frequency domain and time domain is that it has capabilityto handle both, stationary and non-stationary vibration signals.

III. Machine Learning Architecture

There are different types of classification methods for bearing conditions. The classification can be done for the fault type/position (healthy condition or fault on the inner ring, outer ring, balls or the cage) or the fault severity. Fault severity classification is possible for the fault size or for fault categories. In addition, the classification types can also be used together. Apart from different conditions, the bearing fault detection is also performed using different source of data input. Most of the research works use the vibration data measured from the bearing. A few of the research works are based on the stator current input data to monitor the condition of the bearing.





1. Neural Network Model:

This section presents an overview of available approaches for ANN techniques used for bearing classification. Table 2 shows a list of relevant work in this area. The presented technique is Deep Belief Network (DBN), Wavelet Neural Network, unsupervised networks (K nearset), CNN applied in different shapes.

Authors	Dataset	Signal analysis methods	Diagnosis methods	Success rate (%)	Classification type	Key points
Lei et al. (2011)	Inhouse (450 samples; 9 conditions)	Time domain: 4 features Frequency domain: 7 features Time-frequency domain: EEMD	WNN (ANN with a wavelet activation function)	91,65–100	10 positions [compound faults]	This paper analyzes compound faults e.g. inner and outer ring at the same time. The major component of the proposed approach is the data pre-processing.
Cococcioni et al. (2013)	Inhouse (1734 samples; 10 bearings)	Frequency domain: FFT	QDC,LDCMLPRBF	99	5 positions [2 inner, rolling element, cage, healthy]	They have integrated noise reduction, but the used features are not explained.
					3 sizes [light, medium, high]	Identification of specific type of the fault Robust model Excellent accuracy
Delgado et al. (2013)	CWRU	Time domain: 15 features.	MLP	95	10 positions and sizes [inner, outer, rolling element, healthy; 3 fault sizes for each]	Two layered mechanism. The first layer is for the fault type and the second layer is for the fault sizes They focus more on the importance of feature selection in dataset
Muruganatham et al. (2013)	Dataset 1: CWRU	Feature selection with the help of an LDA. Feature extraction with a CCA Time domain: singular value decomposition (SVD)	BPNN	95	4 positions [inner, outer, rolling element, healthy]	This paper verifies their solution on two different models for two different bearing datasets.
	Dataset 2: University of New South Wales					Relatively easier implementation Effective noise tolerance
Gan et al. (2016)	CWRU	Time-frequency domain: energy spectrum features out of WPT	Hierarchical diagnosis network (two layered DBN) compared to SVM and BPNN	99	4 positions [inner, outer, rolling element, healthy]	Classification based on two layers. First layer for fault type and second layer for fault size.
					- 3 sizes [light, medium, high]	Higher complexity Useful for multi-stage diagnosis Better performance even with the noise
Chen et al. (2017)	Universidad Politécnica Salesiana, Ecuador	Time domain: 7 features	Deep Boltzmann Machines; Deep Belief Networks; Stacked Autoencoders	99	4 positions [inner, outer, rolling element, healthy]	The best results are achieved by a combination of frequency and time-frequency domain features.
		frequency domain: RMS time-frequency domain: energy spectrum (WPT)				They also showed that deeper neural networks not always lead to better results.
Lu et al. (2017b)	CWRU	Time domain: Raw data	SDA	91.79 – 95.58 depending on the selected datasets	4 positions [inner, outer, rolling element, healthy]	Can reconstruct the original de-noised signal. Used one fault size to classify other fault sizes
Shao et al. (2017)	Dataset 1: Gearbox data (88 training samples and 40 testing samples for each of the 5 fault types)	Time domain: Raw data	Deep autoencoder optimized by an artificial fish swarm algorithm	Dataset 1: 94.05	Dataset 1: 5 types [healthy, eccentric fault, abrasion fault, misalignment fault, spalling fault]	Sample length matters: longer samples lead to better results (~2.2 %).
	Dataset 2:LLocomotive bearing fault (512 samples for each of the 9 fault positions)			Dataset 2:87.8	Dataset 2: 10 positions [compound faults]	This mechanism works for small bearings (gearbox) and large bearings (locomotives).

Authors	Dataset	Signal analysis methods	Diagnosis methods	Success rate (%)	Classification type	Key points
Guo et al. (2016)	CWRU	Time domain: CNN: Raw data; SVRM: 9 features	Two-tier detection method based on. First tier fault type and second tier fault size. The ADCNN which has an adaptive learning rate	99.3	10 positions and sizes [inner, outer, rolling element, healthy; 3 fault sizes for each]	Fault pattern and size recognition Comparison with other well-known methods Auto feature detection
Ding and He (2017)	CWRU	Time-frequency domain: WPI	ConvNet	96.8	6 positions and sizes [2 sizes inner, 1 size outer, 2 sizes rolling element, healthy]	For pushing the results, they have chosen combinations which are easy to detect (e.g. only large faults were selected). Results compared with other approaches Excellent multi-class classification achieved
Jing et al. (2017)	Dataset 1: 2009 PHM data	Time domain: 8 features	CNN	Time-domainfeatures: 48.32	6 types [mixture of gear, bearing and shaft faults]	For classification of the bearing dataset the best results were achieved by using frequency domain features.
	Dataset 2:planetary gearbox	Frequency domain: RMS of 32 frequency bands		Frequency domain features:99.33		There classification approach combines bearing faults with other gearbox faults e.g. shaft faults.
		Time-frequency domain: 5 wavelet features		Time-frequency domain features: 93.43		Adaptive feature learning Compared to other approaches
Lu et al. (2017a)	Dataset 1: CWRU Dataset 2: Qian-Peng test rig	Time domain:-unknown	CNN Compared to SDA, SVM, SR	CNN: 90.8 SDA: 91.4	4 positions [inner, outer, rolling element, healthy]	In some test cases additional noise is added to simulate a process in a real environment.
		Frequency domain: unknown		CVM: 64.7 SR: 65.9		Novel feature representation
You et al. (2017)	CWRU	Time domain: Raw data	CNN based feature extraction SVR based classification	96.0	4 positions [inner, outer, rolling element, healthy]	In the CNN the last layer has been replaced with SVR. Hybrid model Superior accuracy achieved over conventional CNN and SVR
Wen et al. (2017)	CWRU	Time domain: Raw data	CNN (LeNet 5) with 8 layers	99.481	10 positions and sizes [inner, outer, rolling element, healthy; 3 fault sizes for each]	Focus is on the process of creating two-dimensional images out of the vibration data Higher accuracy than other techniques Better pre-processing of raw data

Abbreviations: ADCNN = Adaptive Deep Convolutional Neural Network; ConvNet = a deep convolutional network; SDA = Stacked Denoising Autoencoder; SR = Softmax Regression; SVR = Support Vector Regression; SVRM = Support Vector Regression Machine; WPI = Wavelet Packet Energy Image

Figure 1 . Relevant works for classification of bearings using ANN.

2. Support Vector machine:

Authors	Dataset	Signal analysis methods	Diagnosis methods	Success rate (%)	Classification type	Key points
Wang et al. (2012a)	CWRU	Time-frequency domain:EMD with SVD an AR	Improved HSSMC-SVM	96.85	8 positions and sizes [healthy, 4 sizes inner, 3 sizes outer]	Advantage of this method is that no frequencies have to be selected. The features are extracted with an SVD or an AR. SVD has better results.
Du et al. (2014)	CWRU	Time domain: Multifractal features	SVM	94.6	4 positions [inner, outer, rolling element, healthy]	A combination of time domain features and wavelet features delivers the best results.
		Time-frequency domain: 8 WPE features				Discarding redundant features
Saidi et al. (2015)	CWRU	Frequency domain: 8 bi-spectrum features	PCA and SVM-OAA	96.98	4 positions [inner, outer, rolling element, healthy]	Higher accuracy achieved by using only relevant features Method can be used for real-time analysis. Utilization of non-linear features The characteristics of raw data have been excellently exploited in order to extract and use the relevant features
Zhang et al. (2015)	CWRU	Time domain:PE	SVM optimized by inter-cluster distance (ICD)	97.91	4 positions[inner, outer, rolling element, healthy]	Not all classification types in one setup e.g. one setup has all datasets with the same load.
		Time-frequency domain: EEMD				Hybrid model
Li et al. (2016a)	CWRU	Frequency domain: LMD: PF; IMFE	ISVM-BT	96.88	8 positions and sizes [healthy, 3 inner, 2 outer, 2 rolling elements]	Detection of fault type and severity This approach also uses de-noising. Relatively complex implementation Multiple steps in data pre-processing
Li et al. (2016b)	CWRU	Frequency domain: LMD; PF; HFE	ISVM-BT	97.8	8 positions and sizes [healthy, 3 inner, 2 outer, 2 rolling elements]	Same approach as used by Li et al. (2016a). They only replaced IMFE through HFE for better results.
Kang et al. (2017)	CWRU	Time domain: 7 features	SVM	Setup 1: 99.4	4 positions [inner, outer, rolling element, healthy]	Feature reduction with the LLE algorithm gives the best result in this setup.
		Frequency domain: 16 features Time-frequency domain: EEMD	Feature reduction with KPCA, LLE, LPP and LTSA	Setup 2: 91.0		Time and frequency domain features were extracted out of the IMFs of the EEMD.

Abbreviations: AR = Autoregressive model; EEMD = Ensemble Empirical Mode Decomposition; F-ANFIS = Frequency-domain feature based Adaptive Neuro Fuzzy Inference System; F-SVM = Fractal dimensions features based SVM; HFE=Hierarchical Fuzzy Entropy; HSSMC-SVM= Hyper-Sphere-Structured Multi-Class Support Vector Machine; IMFE = Improved Multiscale Fuzzy Entropy; ISVM-BT = Improved Support Vector Machine based Binary Tree; IWPE-SVM = Improved Wavelet Package Energy features based SVM; KPCA = Kernel Principal Component Analysis; LLE = Locally Linear Embedding projection; LMD = Local Mean Decomposition; LPP = Locality Preserving Projection; LS = Laplacian Score; LTSA = Local Tangent Space Alignment; M-FAM = Modified Fuzzy ARTMAP; PCA = Principal Component Analysis; PF = Product Functions; SVD = Singular-Value Decomposition; SVM-OAA = Support Vector Machine – One Against All

Figure 2. Relevant works for classification of bearings using SVMs.

3. Hidden Markov Model:

Authors	Dataset	Signal analysis methods	Diagnosis methods	Success rate (%)	Classification type	Key points
Yuwono et al. (2016)	CWRU	Time-frequency domain: Wavelet Kurtogram- ; Cepstrum liftering	Combination of SRCE an HMM	97.35	6 positions [inner, 3 outer, rolling elements,	Fault frequencies are used during feature extraction. Complex signal pre-processing
Zhou et al. (2016a)	Dataset 1: Inhouse (20 samples per fault position)	Time domain: - 10 different features	CHMM	Dataset 1: 100.0	Dataset 1:- 4 positions [inner, outer, rolling element, healthy]	The presented method is usable for severity and fault type classification.
	Dataset 2: HBRC (4 bearings; 20 samples for each class)	Frequency domain:Envelope energy Time-frequency domain: WPT	Dimensionality reduction of total 21 features with NCA	Dataset 2: 95	Dataset 2:- 4 severities [healthy, early fault, degraded, failure]	Discards the redundant features Improved results by fusing data from different channels
Zhou et al. (2016b)	Dataset 1: Simulated data with strong noise	Time domain: SIDL	HMM	96.25	4 positions [inner, outer, rolling element, healthy]	Detection based on the double impulse which appears when entering and leaving the faulty area of a bearing. Even possible with strong noise.
	Dataset 2: Nuclear Engineering Lab, Toshiba Corp.	Time-frequency domain: WPE				The proposed model relies heavily on the extracted features and their quality
	Dataset 3: Inhouse (30 samples for each fault position)					Adaptive feature extraction

Abbreviations: CHMM = Coupled HMM; HBRC=Hangzhou Bearing Test & Research Centre; NCA = Neighborhood Component Analysis; SIDL = Shift Invariant Dictionary Learning; SRCE = Swarm Rapid Centroid Estimation; WPT = Wavelet Packet Transformation

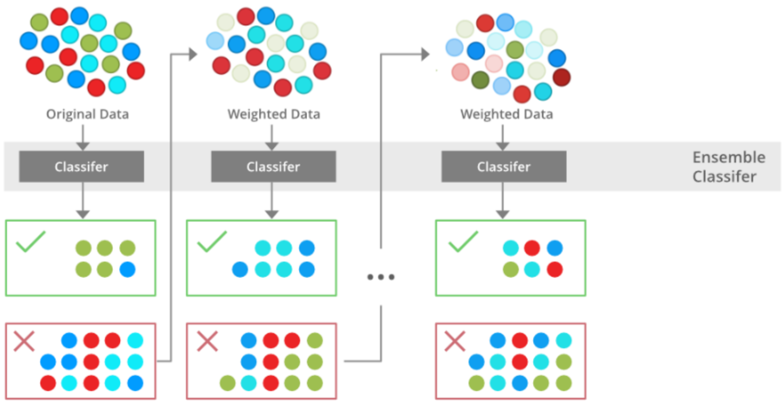
Figure 3 . Relevant works for classification of bearings using HMMs

4. State of Art Method:

Due to many competition on kaggle. the state of art machine learning architecture nowadays is DWT + GradientBoosting approach (XGboost).

Why XGboost:

XGBoost gained significant favor in the last few years as a result of helping individuals and teams win virtually every Kaggle structured data competition. In these competitions, companies and researchers post data after which statisticians and data miners compete to produce the best models for predicting and describing the data.



Boosting

Boosting is an ensemble modelling, technique that attempts to build a strong classifier from the number of weak classifiers. It is done by building a model by using weak models in series. Firstly, a model is built from the training data. Then the second model is built which tries to correct the errors present in the first model. This procedure is continued and models are added until either the complete training data set is predicted correctly or the maximum number of models are added.

```
training Nearest Neighbors ...
Score ofNearest Neighbors is: 0.7156105100463679
training Linear SVM ...
Score ofLinear SVM is: 0.5691653786707882
training RBF SVM ...
Score ofRBF SVM is: 0.7187017001545595
training Decision Tree ...
Score ofDecision Tree is: 0.731839258114374
training Random Forest ...
Score ofRandom Forest is: 0.6877897990726429
training Neural Net ...
Score ofNeural Net is: 0.6387171561051005
training AdaBoost ...
Score ofAdaBoost is: 0.597758887171561
training Naive Bayes ...
Score ofNaive Bayes is: 0.5467542503863988
training QDA ...
Score ofQDA is: 0.6796754250386399
Score ofXGBoost is: 0.8423493044822257
training CatGBoost ...
```

Figure 4: Testing Score take from Kaggle paper of FURKAN ÇITIL

IV. Demo Fully-Connected Neural Network on DWT Data

Link code colab demo:

https://colab.research.google.com/drive/1E4u0xTW8_mJM1nGIrl-YP7vW4cn-JdnC?usp=sharing

https://colab.research.google.com/drive/1YlRIUuvkn4r7xUAo40J8CmDoXblAvLYi?usp=sharing

1. The Dataset:

The dataset of Vibration Measurements on a Rotating Shaft at Different Unbalance Strengths

(https://fordatis.fraunhofer.de/handle/fordatis/151.2)

Datasets for 4 differently sized unbalances and for the unbalance-free case were recorded. The vibration data was recorded at a sampling rate of 4096 values per second. Datasets for development (ID "D[0-4]") as well as for evaluation (ID "E[0-4]") are available for each unbalance strength. The rotation speed was varied between approx. 630 and 2330 RPM in the development datasets and between approx. 1060 and 1900 RPM in the evaluation datasets. For each measurement of the development dataset there are approx. 107min of continuous measurement data available, for each measurement of the evaluation dataset 28min. Details of the recorded measurements and the used unbalance strengths are documented

TABLE I
PARAMETERS OF USED DATASETS

ID	Radius [mm]	Mass [g]	Unbalance Factor [mm g]	Number of Samples Development	Number of Samples Evaluation
0D / 0E	-	0	0	6438	1670
1D / 1E	14 ± 0.1	3.281 ± 0.003	45.9 ± 1.4	6434	1673
2D / 2E	18.5 ± 0.1	3.281 ± 0.003	60.7 ± 1.9	6434	1669
3D / 3E	23 ± 0.1	3.281 ± 0.003	75.5 ± 2.3	6430	1672
4D / 4E	23 ± 0.1	6.614 ± 0.007	152.1 ± 2.3	6430	1675

Here is the data that you can see from the first 2 graph above which include the first 10 seconds . Since the tutorial said that the first about 10 seconds are noisy due to the warm-up phase of the measuring device. So the first 50000 samples are skipped and we get 4 graphs below:

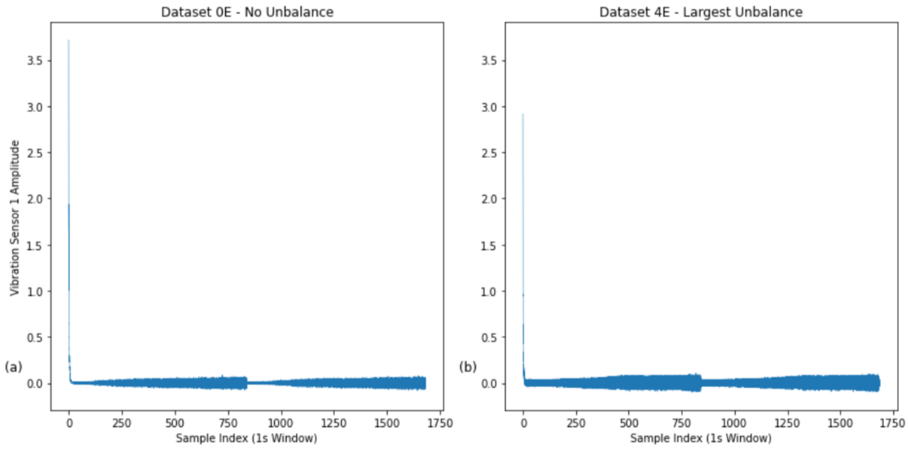


Figure 1: dataset without skipping 50000 samples (x axis scaled to second unit (4096 sample per window) which I also called the index of sample window)

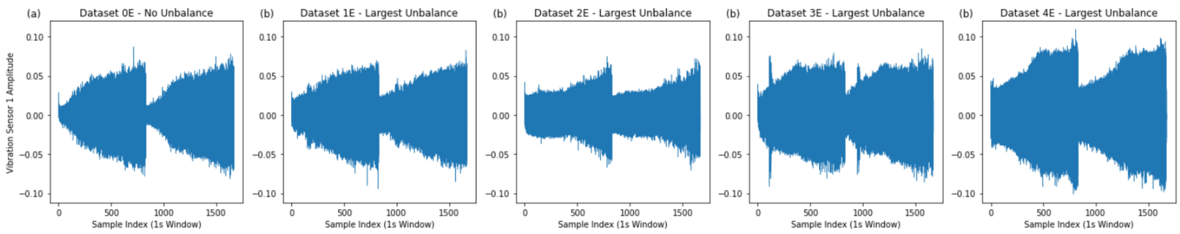


Figure 2: Raw dataset of 5 different sized unballanced (x axis scaled to second unit)

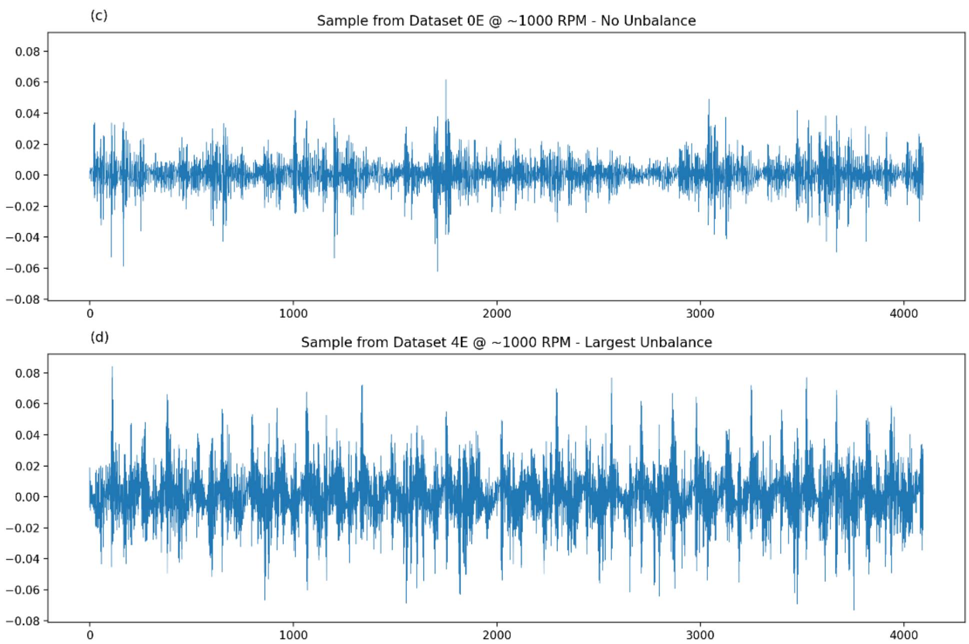


Figure 3: Plotting samples for 1 window

Applied FFT for windows taken by moving Hann Window with 2^{12} samples, fs =4096 Hz , overlaping = 2^{11} samples , we got:

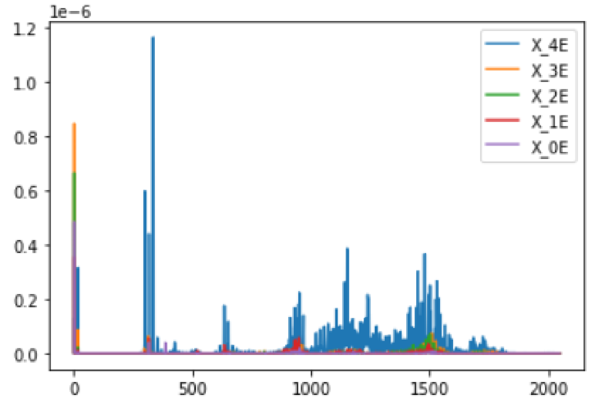


Figure 4: take absolute value of Fast Fourier transform for the first Window

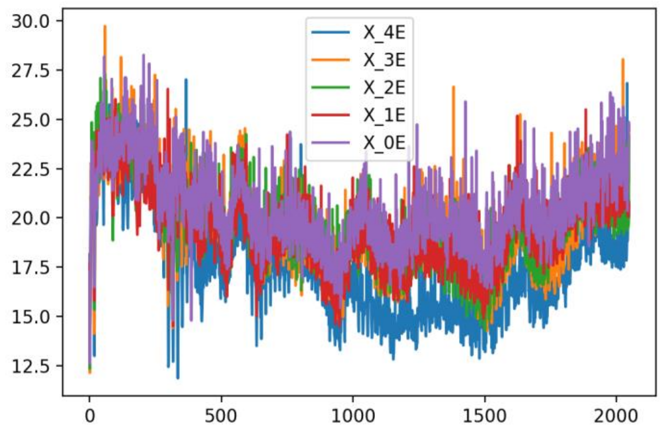


Figure 5: Apply semilogy for FFT data we got the data

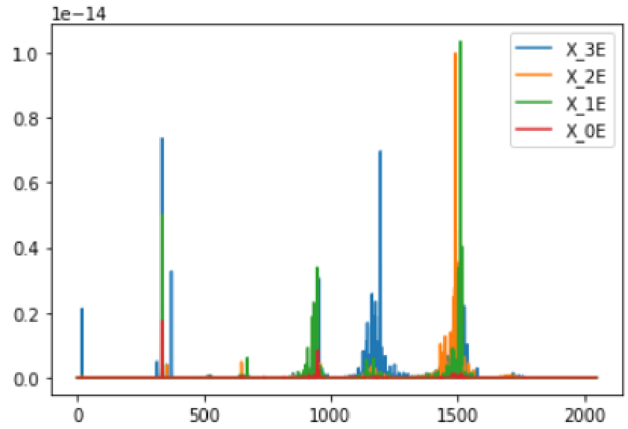
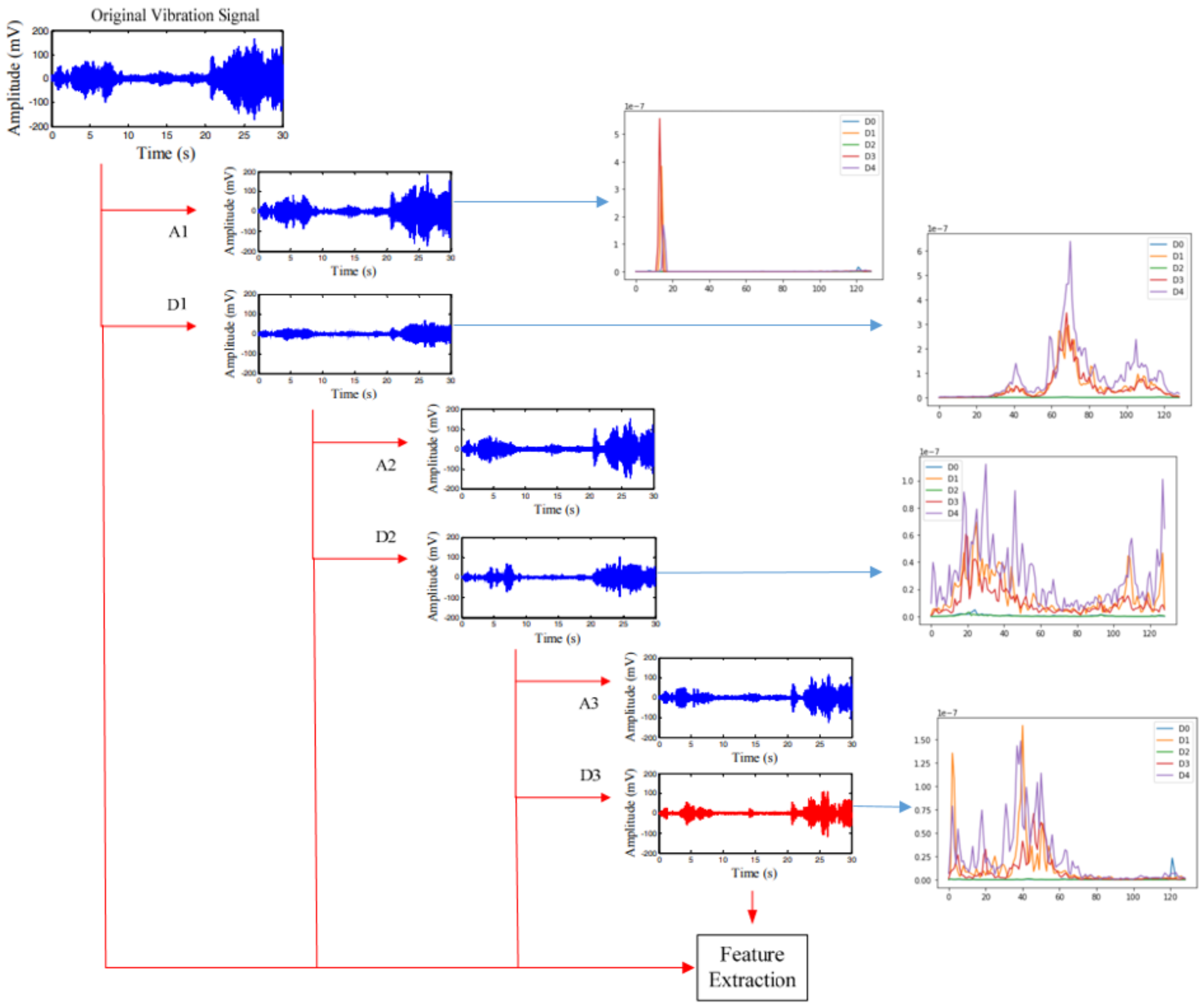


Figure 6: plotting PSD after FFT for 3 types of vibraiton

Chosing Wavelet transform and then calculate the PSD to get better feature extraction.

Using the approaching feature extraction which inspired from the gragh in a paper name: A guide for using the Wavelet Transform in Machine Learning via https://ataspinar.com/

Using wavelet transform and wavelet decomposition for discrete dataset to filt decomposite the frequency from the total raw times domain signal.



This is the frequency in each time step (D0= 15s, D1 =465s, D3 =915s, D4= 1350s) by using Wavelet decomposition to filter the raw data Window and apply FFT to get the spectrum of frequency in the wave shape signal.

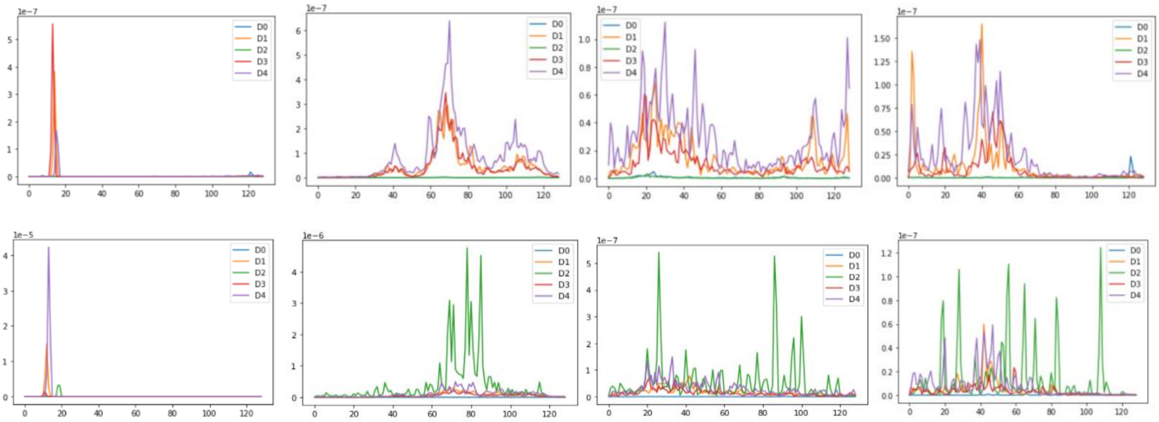


Figure 7: 4 couples of graph (above-below) represent for 3 levels of composite wave signal (A ,D1 ,D2 ,D3 from left to right respectively).

In the feature extraction step, I apply many scaler method for each wavelet window and then horizontal stack all arrays of windows wavelet together to get the better information features. Beacause the date is just from 1 dimenstion (one sensor 'Vibration_1' so that the data graph is some times look the same in some time steps.

After extracting raw data, we get the features data for train and validation like this:

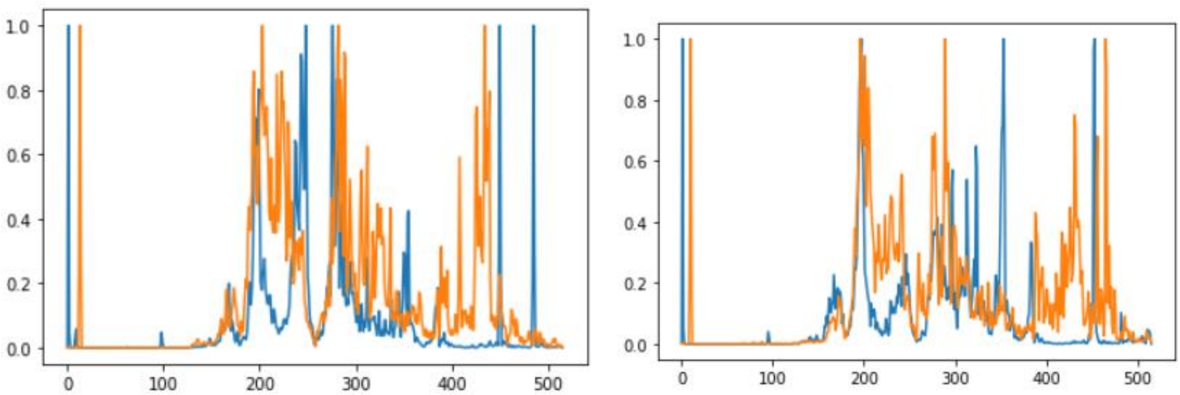
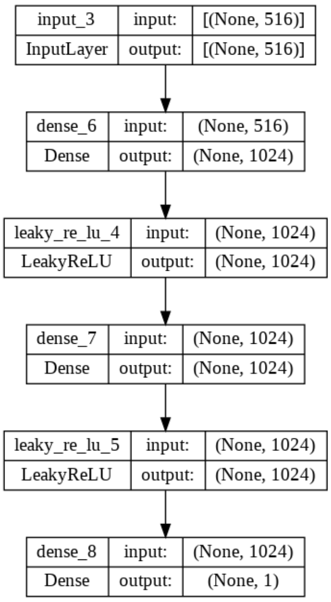


Figure 8: the graph include 2 signal of time (0s) and (1700s)

2. Machine learning model:

In this project to due with the simple 1D data, I chose the simple architecture of neural network (include 2 layers of fully-connected complie with LeakyReLU layers). In this demo project, I just use data from one sensor 'Vibration_1'. For the better performace, If the computer can adapt to large memory job, I suggest use data from all sensor you have and use Xgboost instead.



3. Result:

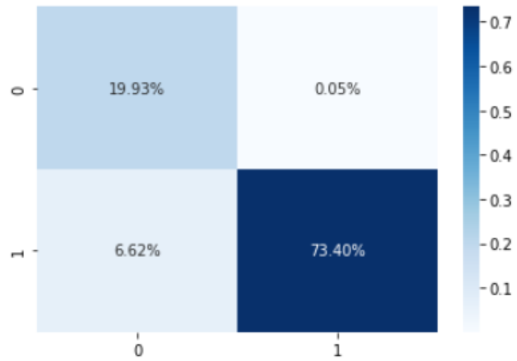
```
Epoch 37: saving model to /content/drive/MyDrive/data/training_2/cp-0037.ckpt
503/503 [==============================] - 21s 42ms/step - loss: 2.5355e-09 - accuracy: 1.0000 - val_loss: 0.6807 - val_accuracy: 0.9307
Epoch 38/100
 35/503 [=>............................] - ETA: 17s - loss: 1.8130e-09 - accuracy: 1.0000
Epoch 38: saving model to /content/drive/MyDrive/data/training_2/cp-0038.ckpt
503/503 [==============================] - 22s 44ms/step - loss: 2.1752e-09 - accuracy: 1.0000 - val_loss: 0.6547 - val_accuracy: 0.9337
Epoch 39/100
 36/503 [=>............................] - ETA: 18s - loss: 1.1441e-09 - accuracy: 1.0000
Epoch 39: saving model to /content/drive/MyDrive/data/training_2/cp-0039.ckpt
503/503 [==============================] - 21s 41ms/step - loss: 1.7018e-09 - accuracy: 1.0000 - val_loss: 0.6669 - val_accuracy: 0.9329
Epoch 40/100
 37/503 [=>............................] - ETA: 17s - loss: 1.9503e-09 - accuracy: 1.0000
Epoch 40: saving model to /content/drive/MyDrive/data/training_2/cp-0040.ckpt
503/503 [==============================] - 21s 41ms/step - loss: 1.3975e-09 - accuracy: 1.0000 - val_loss: 0.6854 - val_accuracy: 0.9316
Epoch 41/100
 39/503 [=>............................] - ETA: 17s - loss: 1.6830e-09 - accuracy: 1.0000
Epoch 41: saving model to /content/drive/MyDrive/data/training_2/cp-0041.ckpt
503/503 [==============================] - 21s 41ms/step - loss: 1.1740e-09 - accuracy: 1.0000 - val_loss: 0.6726 - val_accuracy: 0.9332
Epoch 42/100
```

Here is the prediction result:

```
model.evaluate(x=X_val,y=y_val)
```

```
523/523 [==============================] - 3s 6ms/step - loss: 0.6985 - accuracy: 0.9333
[0.6985066533088684, 0.9333453178405762]
```

Conffusion Matrix:



Can see that the model work well on classify balanced, and not good too much on classify.

V. Conclusion:

In this report, an attempt has been made to summarize the recent trends in research on vibration analysis of defects in rolling element bearing and techniques for fault detection in time, frequency and time frequency domain. Researchers have developed several techniques for measuring vibration and are still attempting to improve signalprocessing techniques. The time domain method uses scalar indicators like RMS, crest factor, kurtosis for detection of defect, but this technique cannot identify defect location. The advantage of vibration measurement in frequency domain is that it can identify defects as well as the defect location. The high-frequency resonance technique is widely used technique in the frequency domain but has a limitation that when the damage is advanced, the defect frequency may be submerged in the rising background level of the spectrum. The time-frequency domain approach is an effective signal processing technique for both stationary and non-stationary vibration signals. The wavelet transform is widely used technique in the time - frequency domain because it is capable of extracting weak signals for which FFT is ineffective. In the other hand, there are different types of classification methods for bearing conditions. The classification can be done for the fault type/position, but it's better to compile many way of classifying together to get the best result. This report also notice about the state of art method for vibration analysis is combine Discrete Wavelet Transform + GradientBoosting approach (XGboost).

VI. Reference:

[1] https://www.kaggle.com/code/furkancitil/nasa-bearing-dataset-supervised-learning/notebook#Conclusion

[2] Fault detection analysis in rolling element bearing: A review Pankaj Guptaa * Department of Mechanical Engineering, Maulana Azad National Institute of Technology, Bhopal (M.P.), India , M. K Pradhanb (link: https://scihub.se/https://www.sciencedirect.com/science/article/pii/S221478531730250X)

[3] Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics Hai Qiua, Jay Leea , Jing Linb , Gang Yuc (link: https://scihub.se/https://www.sciencedirect.com/science/article/abs/pii/S0022460X0500221X)

[4] https://blog.endaq.com/top-vibration-metrics-to-monitor-how-to-calculate-them#summary

[5] Machine Learning-Based Unbalance Detection of a Rotating Shaft Using Vibration Data Oliver Mey, Willi Neudeck, Andre Schneider and Olaf Enge-Rosenblatt (linkhttps://ieeexplore.ieee.org/document/9212000)

[6] A guide for using the Wavelet Transform in Machine Learning (link: https://ataspinar.com/2018/12/21/a-guide-for-using-the-wavelet-transform-in-machine-learning/)

[7] https://blog.endaq.com/vibration-analysis-fft-psd-and-spectrogram

[8] Vibration Analysis: FFT, PSD, and Spectrogram Basics (get from : https://github.com/Charlie5DH/PredictiveMaintenance-and-Vibration-Resources)

[9] Time and frequency domain methods for quantifying common modulation of motor unit firing patterns [link: https://jneuroengrehab.biomedcentral.com/articles/10.1186/1743-0003-1-2)

[10] A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing (https://www.mdpi.com/2075-1702/5/4/21)

[11] A survey of machine-learning techniques for condition monitoring and predictivemaintenance of bearings in grinding machines Sebastian Schwendemann∗, ZubairAmjad, Axel Sikora

[12] https://www.geeksforgeeks.org/xgboost/