الجمهورية الديمقراطية الشعبية الجزائرية

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique
Ecole Supérieure des Sciences Appliquées
d'Alger

وزارة التعليم العالي والبحث العلمي
المدرسة العليا في العلوم التطبيقية بالجزائر

Département du second cycle

Mémoire de Fin d'Etudes

En vue de l'obtention du diplôme de MASTER

Filière **: Electrotechnique**

Spécialité : **Traction électrique**

Thème :

Theoretical Study of Rolling Bearing Defect Condition Monitoring Techniques

Présenté par: Bayou Nada

Encadré (e) par : Deboucha Abdelhakim
Co-encadré(e) par : Ounnas Badreddine

Soutenu le : 20/09/2021

Devant le jury composé de :

Dr Hamache Ammar	**Président**
Dr Bensenouci Ahmed	**Examinateur**
Dr Deboucha Abdelhakim	**Encadreur**
Dr Ounnas Badreddine	**Co-Encadreur**

Binôme N°: 21/Master /TR/ 2021

Abstract

Rolling bearing is an important part that is often used in rotating machinery.Rolling bearings faults are one of the main causes of breakdown of these machines. Therefore, the fault diagnosis of rolling bearing is very important to guarantee the production efficiency and plant safety. This document presents different techniques for rolling bearing diagnosis among that the frequency analysis include enevelope analysis using (Fast fourier tronform) FFT and Singular value decomposition (SVD) , In the category of time-domain analysis technique there are STFT, Continuous wavelet transform(CWT), Spectral kurtosis,Time domain include EMD decomposition. In addition A brief introduction of different AI algorithms is presented including the following methods: k-nearest neighbour K-NN), naive Bayes (NB), support vector machine(SVM) , artificial neural network (ANN) and fuzzy neural network. Finally deep learining methods include Convolutional neural network (CNN)and Cyclic Spectral Coherence (CSCoh).

Key words: Bearings fault, fault diagnosis FFT, SVD, STFT, CWT, EMD ,ANN, NB, K-NN, SVM, CNN, CSCoh

Résumé:

Le roulement à rouleaux est une pièce importante qui est le plus souvent utilisée dans les machines tournantes Les défauts des roulements à rouleaux sont l'une des principales causes de panne des machines tournantes. Par conséquent, le diagnostic des défauts des roulements est très important pour garantir l'efficacité de la production et la sécurité de l'usine. Cet étude présente différentes techniques pour le diagnostic des roulements, parmi lesquelles l'analyse de fréquence inclut l'analyse d'enveloppe utilisant la FFT (Fast fourier tronform) et la décomposition en valeur singulière (SVD). Dans la catégorie des techniques d'analyse du domaine temporel, il y a STFT, Transformation en ondelettes continue (CWT), Kurtosis spectral Le domaine temporel inclut la décomposition EMD. De plus, une brève introduction des différents algorithmes d’intelligence artificielle est présentée, y compris les méthodes suivantes : k-voisin le plus proche (KNN), Bayes naïf (NB), machine à vecteurs de support (SVM), réseau de neurones artificiels (ANN) et réseau de neurones flous. Enfin, les méthodes d'apprentissage en profondeur incluent le réseau de neurones convolutifs. (CNN) et Cohérence Spectrale Cyclique (CSCoh).

Mots clés : Défaut de roulement, diagnostic de défaut FFT, SVD, STFT, CWT, EMD,ANN, NB, K-NN, SVM, CNN, CSCoh

الملخص:

تعتبر المحامل الدوارة جزءًا مهمًا يستخدم غالبًا في الآلات اللاتزامنية تعتبر أعطال المحامل الدوارة أحد الأسباب الرئيسية لانهيار الآلات اللاتزامنية ، علاوة على ذلك ، قد تكون مشكلة خطيرة أيضًا بسبب الفشل. لذلك ، فإن تشخيص خطأ المحمل الدوراني مهم جدًا لضمان كفاءة الإنتاج وسلامة المصنع. تقدم هذه الدراسة تقنيات مختلفة لتشخيص المحامل المتدحرجة من بينها تحليل التردد الذي يشمل تحليل التطور باستخدام (FFT) والتحليل الفردي للقيمة (SVD). في فئة تقنية تحليل المجال الزمني هناك STFT. التحويل الموجي المستمر (CWT) ،التفرطح الطيفي Spectral kurtosis ، يشمل المجال الزمني تحليل EMD. بالإضافة إلى ذلك ، يتم تقديم مقدمة موجزة عن خوارزميات الذكاء الاصطناعي المختلفة بما في ذلك الطرق التالية: k- الجار الأقرب (K-NN) , (Naive Bayes (NB ، آلة ناقلات الدعم (SVM) ، الشبكة العصبية الاصطناعية (ANN) والشبكة العصبية الضبابية أخيرًا تقنيات التعلم العميق تشمل طرق التعلم الشبكة العصبية التلافيفية (CNN) والتماسك الطيفي الدوري (CSCoh).

الكلمات المفتاحية: عطل المحامل ، تشخيص الأعطال,FFT ، SVD ، STFT ، CWT ، ANN EMD ، NB ، K-NN، SVM ، CNN ، CSCoh

Aknowledgements

In the name of Allah, the Most Gracious and the Most Merciful. All praises to Allah and His blessing for the completion of this thesis. I thank God for all the opportunities, trials and strength that have been showered on me to finish writing the thesis

First and foremost, I would like to sincerely thank my supervisor Dr. DEBOUCHA for his guidance, understanding, patience and most importantly, he has provided positive encouragement and a warm spirit to finish this thesis. It has been a great pleasure and honor to have him as my supervisor.

My deepest gratitude goes to all of my family members. It would not be possible to write this thesis without the support from them. I would like to thank my father , my mother my brothers and sisters, for their motivation, prayers and their sincere help during my studies .

Finally my heartfelt thanks go to all my teachers and friends for their help and support during my studies.

Table of contents

List of Figures

LIST OF SYMBOLS

$\Psi(t)$:Wavelet basic signal

$X(t)$: Input signal

τ: Translation factor

s: Scal factor.

r_n: Central tendency of signal $x(t)$

A: Real matrix

U,W: Orthonormal matrices.

D: Diagonal matrice.

f: Activation function

H1, H2: Hyper planes

X_i,Y: Booleen variable

k_j: Convolution kernel

E: The statistic mean

T: Cyclic period

I Introduction

Electric motors have been used in various industrial sectors to convert electrical energy to mechanical energy. Induction motors known as workhorse of modern industries are subjected to some undesirable stresses during their operating lifetime, causing some faults to develop leading to failures(P.konar,et al,2011). Motors fail due to reasons such as shorting of stator winding, bent shaft, broken rotor bar, and motor bearing failure. Different sources have demonstrated that motor bearing failure is the top reason that electric motor fails, since motor reliability studies shows that bearing faults accounts for 44% of the faults occurring in an induction motor(P.Konar, et al,2011). Bearing failures can be caused by several factors, such as incorrect design or installation, acid corrosion, poor lubrication, humidity and water are commonly present and cause mainly major deterioration factors namely corrosion and contamination(R. RUBINI,et al,2001). Bearing defects can be classified as distributed or local. The distributed defects are surface roughness, waviness, misaligned races, and off-size rolling elements. The local defects include cracks, pits, and spalls on the rolling surfaces (S. Prabhakar,et al 2002). Machine health monitoring is of great importance in modern industry. Failure of these machines could cause great economical loss, and sometimes poses threats to the people who work with the machines, If faults can be detected early, the failure frequency and severity can be reduced as a result of optimized maintenance. Thus, finding an efficient and reliable fault diagnostic technique, especially for induction motors, is extremely important due to widespread use of automation and consequent reduction in direct man–machine interface to supervise the system operation. Detection of rolling element-bearing faults has been gaining importance in recent years because of its detrimental effect on the reliability of the induction machines. Vibration analysis has been used in rotating machines fault diagnosis for decades. Each fault in a rotating machine produces vibrations with distinctive characteristics that can be measured and compared with reference ones in order to perform the fault detection and diagnosis (P.Konar, et al.2011). The traditional method of fault monitoring sach as vibrations-whether felt by hand or measurement with a frequency analyzer, abnormal noises, displacement of rotational centerline,…etc is carried out by experts, they are well versed in understanding the sound patterns of equipment, but analysis based on human perception can sometimes fail in analyzing the exact nature of the signal. This analysis can be done effectively using sensor technology and processing techniques . Acoustic sensors and vibration sensors are used to monitor the condition of the bearing. The analysis of the condition of the bearings in case of acoustics is better suited when the direct translation of vibrations is not effective, that is to say if the machine is located at a distance or immersed in the fluid. The same is true of vibration when the operating environment is too noisy, the addition and multiplication of signal vectors can mislead the information content in the actual signal. Appropriate signal processing techniques must be chosen to extract the signal from the noise . The signal processing can be done through different techniques i.e., Fast Fourier Transform (FFT), Short Time Fourier Transform (STFT), Wavelet Transform and Hilbert Huang Transform (HHT) . FFT can only be used for the signal of stationary in nature, but it cannot be applied to transient and non-stationary signals. STFT can be used to analyze the non-stationary signal, but selection of window size as per the frequency requirement is tough to decide for random signals. Wavelet transform is better option than FFT, variable window size allows the possibility to extract both low frequency as well as high frequency information as per requirement, but the selection of basis function affect the analysis of transient signals. Empirical mode decomposition can be applied to non-stationary and transient signals. The signal obtained from the sensors are decomposed in to number of IMFs using empirical mode decomposition (EMD) technique. Later on FFT is applied to the IMFs to get beck the frequency (S. Mohanty, et al,2013). Although kurtosis based on temporal signals is effective under some conditions, its performance is low in the presence of a low signal-to-noise ratio and non-Gaussian noise. It is logical to measure kurtosis in the frequency domain. When a bearing is healthy, its envelope spectrum is randomly distributed over that of the whole frequency. However, when it has localised faults, bearing fault characteristic frequencies dominate the envelope spectrum(D. Wang, et al,2013) Moreover, SVD applied in bearing fault detection. SVD decompose the vibration signals which is collected by the acceleration, it makes it possible reveal the different characteristic frequencies of the defects which do not appear clearly during a first wavelet transform (X. Milisien, et al, 2006). In recent years AI algorithms for fault diagnosis of rotating machinery have become popular due to their robustness and adaptation capabilities they are widely adopted due to the fast implementation and good classification performance. Also, they do not require full prior physical knowledge, which may be difficult to obtain in practice. Among the various AI algorithms, k-NN, fuzzy neural network,Naive Bayes classifier, SVM and ANN algorithms have been applied most commonly in fault diagnosis(R. Liu, et al,2018). Support Vector Machines (SVMs) have been found to be remarkably effective in many real-world applications. SVMs have been successfully applied in various classification and pattern recognition tasks, but in the area of fault diagnostics they have not been widely studied. SVMs are based on statistical learning theory and they specialize for smaller number of samples (P.Konar, et al,2011). With NN, it is possible to estimate a nonlinear function without requiring a mathematical description of the functional relationship between input and output. The most commonly mentioned advantages for neural networks are their ability to build a model for any nonlinear system, the ability to learn for strongly parallel structures, and the ability to process conflicting or noisy data(S. Samir Hamdani,2012). A fuzzy neural network is used to automatically distinguish the fault types of a bearing by time domain features. Both neural networks and fuzzy systems have some things in common. They can be used for solving a problem if there does not exist any mathematical model of the given problem However, Neural networks can only come into play if the problem is expressed by a sufficient amount of observed examples. These observations are used to train the black box, On the contrary, a fuzzy system demands linguistic rules instead of learning examples as prior knowledge. Furthermore the input and output variables have to be described linguistically. If the knowledge is incomplete, wrong or contradictory, then the fuzzy system must be tuned. Since there is not any formal approach for it, the tuning is performed in a heuristic way. This is usually very time consuming and error-prone.

KNN is a supervised learning algorithm and widely applied in pattern recognition. In comparison with other types of supervised learning algorithms such as neural networks, KNN is a non-parametric method because it makes no assumptions about the relationship between the outcome variable and the predictors and therefore does not involve estimation of parameters of the assumed functions nor does it need a model at the training stage(D.He, et al,2011). In addition, The fault classification has been accomplished with high success rate by using the extracted fault features with the naive Bayes classifier due to its inherent features like robustness to noise, simplicity, and quick convergence even with less training data. The efficacy of the proposed classifier has been verified with both the offline signals and the real-time signals having different fault severity levels (M. K.Saini, et al,2018).

Traditional classification techniques such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and k-NN are widely adopted due to the fast implementation and good classification performance. However, being limited to their shallow architectures, they have difficulty in learning effectively discriminative features from raw high-dimensional inputs. The diagnosis performance relies strongly on the diagnosis expertise and the diagnostic feature quality. In recent years, DL-based intelligent fault diagnosis approaches have received increasing attention. Among the DL algorithms, Convolutional neural network CNNs have been widely used and achieved state-of-the-art diagnosis performance for their strong local feature extraction capability and flexible architectures. (Guo, et al,2016) converted raw data into input matrices and then designed a CNN hierarchical framework with adaptive learning rate to recognize bearing fault patterns and fault sizes. The development of faults in rolling element bearings lead to the generation of repeated impacts due to the passing of rolling elements over the defect. The collected time domain signals usually cannot directly reveal the intrinsic characteristics of bearing health conditions. Valuable information, i.e. periodicities and correlations related to the failure modes, are easily masked by strong background noise and other components. To overcome this weakness, in the preprocessing step, by considering the bearing health conditions and the physical characteristics of the defects, the cyclic spectral analysis is proposed to obtain 2D CSCoh images/maps. CSCoh constructs a frequency-frequency domain map to exhibit the amplitude levels of the modulation frequency of excitations by exploiting the second-order cyclostationary behavior. (Z. Chen, et al,2020).

In this paper we will see some details about all the monitoring technique montioned above usesd for diagnosis of bearing faults.

II Theoretical Study of Rolling Bearing Defect Condition Monitoring Techniques

There are numerous condition monitoring techniques for detecting bearibg fault in which some of them are discussed in this letter:

II.1 The envelope method

The envelope method is a technique using high frequency resonance of the bearing (or sensor). It uses the resonant frequency1 of the bearing to extract the information necessary to determine the presence of the defect and highlights this information in a frequency range normally observed in vibration analysis (0 - 1500 Hz). More precisely, the envelope method uses the modulation of the amplitude of the bearing's resonant frequency , by the frequency of the defect. The envelope method requires a series of processing of the raw time signal before obtaining the result. The first step is the filtration of the raw signal in order to eliminate the unwanted components: this promotes the robustness of the method against noise. Then the envelope is calculated: it is sort of a signal rectification; at this time, we have time information. Finally, using the Fourier transform, we obtain the spectrum of the envelope, which is a graph of the amplitude of vibration as a function of frequency. It is from this spectrum that conclusions will be drawn (X. MILISEN, et al, 2006).

II.1.2 Amplitude Modulation

Amplitude modulation of a frequency is the periodic variation in the amplitude of the signal over time. For example in the event of a fault in the inner race, the resonant frequency is modulated by the frequency of the inner race fault (BPFI). The impact generated by the rolling element on this fault produces a vibration at the resonant frequency. At the precise instant when the impact occurs, the amplitude of the associated vibration is maximum. This amplitude then decreases over time, which is due to the damping of the structure. The impact is a periodic phenomenon, occurring at a frequency characteristic of the fault. Whenever this phenomenon occurs, it generates a vibration at the resonant frequency. Thus, the amplitude of the vibration at the resonant frequency varies with a period equal to the repetition period of the impact, characteristic of the fault: the amplitude is modulated. Using the envelope method, this amplitude modulation can be extracted from the original signal. Indeed, the envelope of a signal modulated in amplitude by another periodic signal is a periodic function of period equal to that of the modulating signal. It will then suffice to interpret the frequency spectrum (obtained by the Fourier Transform) of this envelope in order to remove the impact frequency, specific to the fault. Since this method uses the Fourier Transform, it will only be suitable for stationary signals (signals whose frequency components do not vary over time). In the case where, for example, the speed of rotation of the machine varies over time, this method may then be less effective

II.2 Short-Time Fourier Transform (STFT)

The STFT is a Fourier-based transformation normally used to map a signal into two variables, namely time and frequency of which time is discrete and frequency variable is continuous. STFT is capable to analyze transient signal which varies with time. If in a sequence the time index is fixed then STFT becomes a normal Fourier transform of the sequence. In STFT keeping frequency as fixed quantity interpretation is done as a function of time index. With a particular value of the frequency, interpretation leads to consider the STFT as of linear filtering. Interpretation in terms of linear filtering is useful when only a number of particular frequencies are required to be identified. Then this approach can be used to determine the fundamental frequency and its integer multiple. Another approach is the application of window function to input signal. Advantage of this application is that periodical signal can be analyzed without determining the integer multiple of its periods. There are different types of window functions. Among them Hamming window and Hanning window functions are suitable for harmonic and interharmonic estimation. STFT uses constant-sized window to analyze all frequencies—this is the limitation of this method. This limited window may find it difficult to match the frequency content of the signal which is generally not known prior to the analysis. To overcome, this limited sized window is required to be replaced by a variable-sized window. In WT suitable variable-sized window is use (S.Karmakar, et al,2006).

II.3 Continuous Wavelet Transform (CWT)

The wavelet transform or wavelet analysis is probably the most recent solution to overcome the shortcomings of the traditionally used Fourier transform Wavelets are well suited for approximating data with sharp discontinuities. Motor starting vibration contains numerous non-stationary or transitory characteristics: drift, trends, abrupt changes, and beginnings and ends of events. These characteristics are often the most important part of the signal, and traditional tools like Fourier analysis are not suited for analyzing non-stationary or transitory signals. Fourier analysis is only suitable for steady state analysis consisting of stationary signals – where only the signal's frequency content is needed. When looking at Fourier transform of a signal, it is impossible to tell when a particular event took place, since in transforming to the frequency domain, time information is lost. While in short time Fourier transform (STFT) compromises between time and frequency information can be useful, the drawback is that once a particular size time window is chosen, that window is the same for all frequencies. In wavelet analysis the use of a fully scalable modulated window solves the signal-cutting problem. The window is shifted along the signal and for every position the spectrum is calculated. Then this process is repeated many times with a slightly shorter (or longer) window for every new cycle. In wavelet analysis, the scale plays a special role. Wavelet algorithms process data at different scales or resolutions. If we look at a signal with a large "window", we would notice gross features. Similarly, if we look at a signal with a small "window", we would notice small features. The result in wavelet analysis is to see both the forest and the trees, so to speak Continuous wavelet transform (CWT) is defined as the sum over all time of the signal multiplied by scaled, shifted versions of the wavelet function Ψ (P.Konar, et al, 2011).

$$L_{\Psi} f\left(s,\tau\right) = \int f(t)\psi^{*}_{s,\tau}(t)dt$$

$f(t)$ is decomposed into a set of basis function s, $\psi_{s,\tau}(t)$, called wavelets generated from a single basic wavelet $\psi(t)$, the so called mother wavelet, by scaling and translation:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{|s|}}\psi(\frac{t-\tau}{s})$$

s is a scale factor, τ is the translation factor and the factor $|s|^{-\frac{1}{2}}$ is for energy normalization across the different scale. Scaling a wavelet simply means stretching (or compressing) it. Typical CWT plots for healthy motor and faulty motor with faulted bearing are shown in Fig. 5(a) and (b).

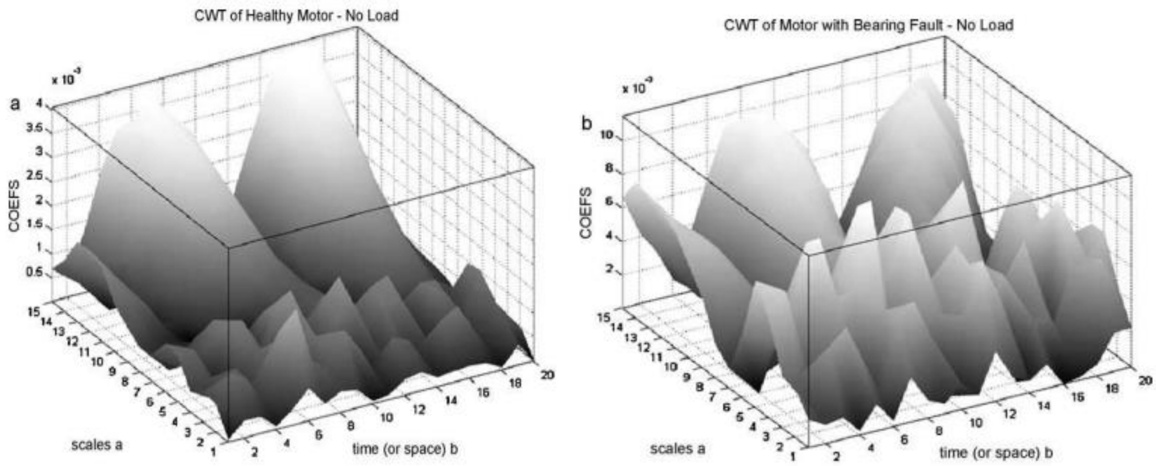

Figure 1:A typical CWT plot of frame vibration signal of (a) healthy, (b) faulty motor(P.Konar, et al, 2011)

II.4 Spectral Kurtosis

Spectral kurtosis (SK) represents a valuable tool for extracting transients buried in noise, it was based on the short time Fourier transform (STFT) and gave a measure of the impulsiveness of a signal as a function of frequency, which makes it very powerful for the diagnostics of rolling element bearings (N. Sawalhi, et al, 2007). When a bearing is healthy, its envelope spectrum is randomly distributed over that of the whole frequency. However, when it has localised faults, bearing fault characteristic frequencies dominate the envelope spectrum . Peak values of bearing fault characteristic frequency and its harmonics can be measured by kurtosis with higher values because kurtosis can be used to measure the protrusion of a signal. Moreover, other faults, such as misalignment, eccentric fault and so on, can also be measured by kurtosis (D. Wang, et al, 2013).

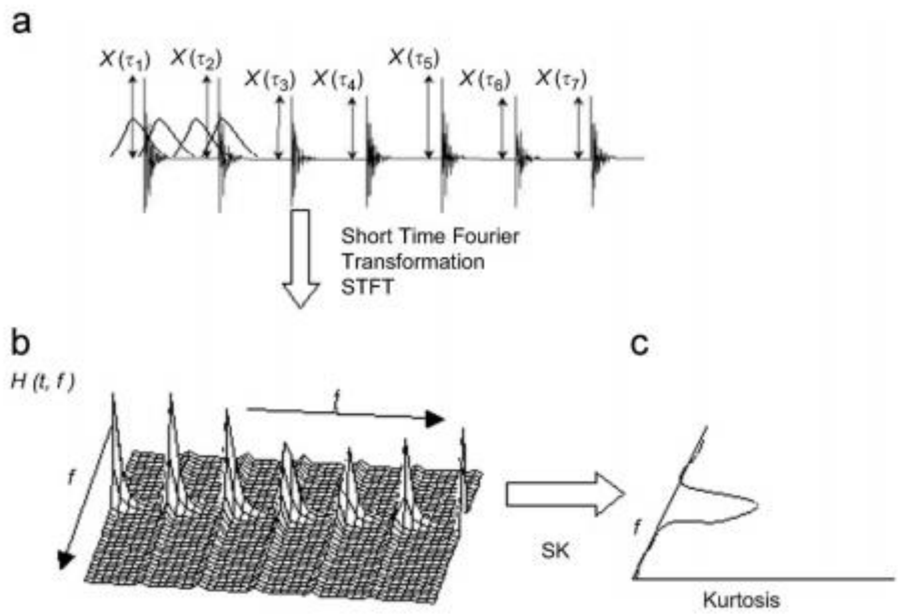

Figure 2:STFT for a simulated bearing fault signal (N. Sawalhi, et al, 2007)

II.5 EMD Method

EMD is a self-adaptive signal processing method that can be applied to nonlinear and non-stationary processes perfectly. The EMD method is developed from the simple assumption that any signal consists of different simple intrinsic modes of oscillations. Each linear or nonlinear mode will have the same number of extrema and zerocrossings. There is only one extremum between successive zero-crossings. Each mode should be independent of the others. In this way, each signal could be decomposed into a number of intrinsic mode functions (IMFs), each of which must satisfy the following definition (Y. Yu,et al,2006)

1. In the whole data set, the number of extrema and the number of zero-crossings must either be equal or differ at most by one.
2. At any point, the mean value of the envelope defined by local maxima and the envelope defined by the local minima is zero.

An IMF represents a simple oscillatory mode compared with the simple harmonic function. With the definition, any signal x(t) can be decomposed as follows (N.E. Huang, et al,1999):

1. Identify all the local extrema, then connect all the local maxima by a cubic spline line as the upper envelope.
2. Repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelopes should cover all the data between them.

3. The mean of upper and low envelope value is designated as m1, and the difference between the signal x(t) and m1 is the first component, h_1, i.e.

$$x(t) - m_1 = h_1 \qquad (1)$$

4. Ideally, if h_1 is an IMF, then h1 is the first component of x(t). (4) If h_1 is not an IMF, h1 is treated as the original signal and (1)–(3) are repeated; then

$$h_1 - m_{11} = h_{11} \qquad (2)$$

After repeated sifting, i.e. up to k times, h_{1k} becomes an IMF, that is

$$h_{1(k-1)} - m_{1k} = h_{1k} \qquad (3)$$

then, it is designated as

$$c_1 = h_{1k} \qquad (4)$$

the first IMF component from the original data. C_1 should contain the finest scale or the shortest period component of the signal.

5. Separating c1 from x(t), we get

$$r_1 = x(t) - c_1 \qquad (5)$$

r_1 is treated as the original data, and by repeating the above processes, the second IMF component c_2 of x(t) could be obtained. Let us repeat the process as described above for n times, then n-IMFs of signal x(t) could be obtained. Then

$$\begin{aligned} r_1 - c_2 &= r_2 \\ &\vdots \\ r_{n-1} - c_n &= r_n \end{aligned} \qquad (6)$$

The decomposition process can be stopped when r_n becomes a monotonic function from which no more IMF can be extracted. By summing up Eqs. (5) and (6), we finally obtain

$$x(t) = \sum_{j=1}^{n} c_j + r_n \qquad (7)$$

Thus, one can achieve a decomposition of the signal into n-empirical modes, and a residue r_n, which is the mean trend of x(t). The IMFs c_1 ,c_2, ... , c_n include different frequency bands ranging from highto low. The frequency components contained in each frequency band are different and they changewith the variation of signal x(t), while r_n represents the central tendency of signal x(t).

First figure bellow shows the vibration acceleration signal of the roller bearing with out-race fault. The decomposed results are given in the second figure. It can be seen from the figures that the signal is decomposed into some IMFs with different time scales by which the characteristics of the signal can be presented in different resolution ratio.

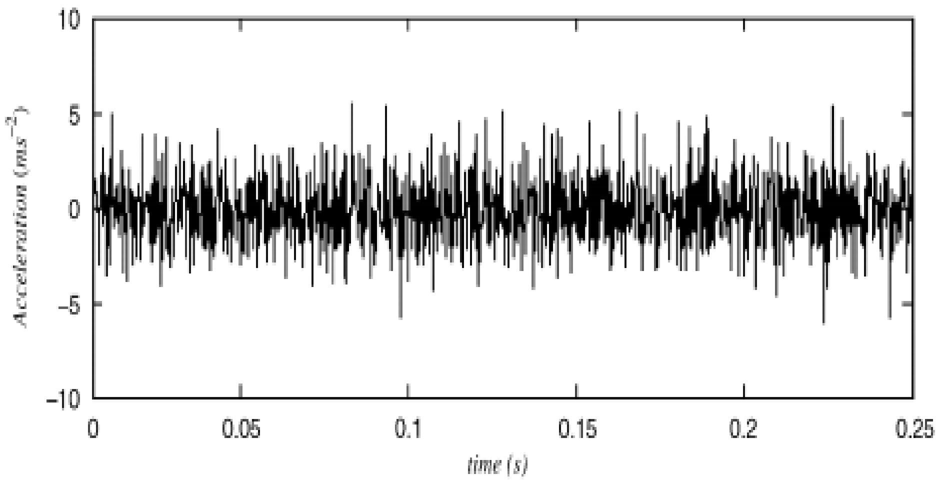

Figure 3:The vibration acceleration signal of the roller bearing with out-race fault

Figure: The vibration acceleration signal of the roller bearing with out-race fault

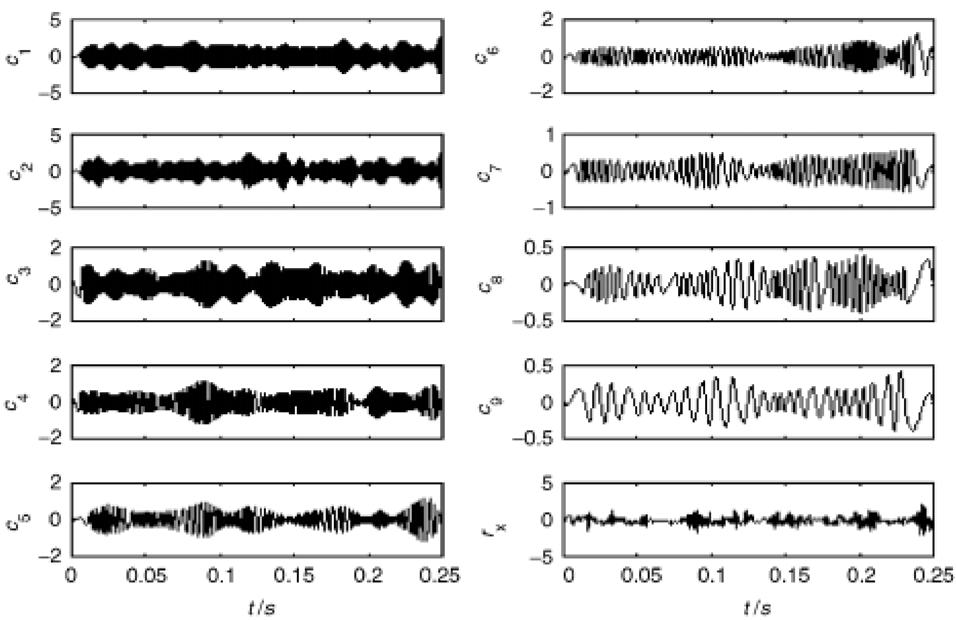

Figure 4: . The EMD decomposed results of vibration signal of the roller bearing with out-race fault

II.6 Singular Value Decomposition (SVD)

Singular value decomposition (SVD) is an important orthogonalization method of matrix decomposition in linear algebra. For a linear correlation matrix of rows or columns, it can be transformed into a linearly independent one by multiplying an orthogonal matrix on its lef and right side, respectively For a real matrix, $A_{m\times n}$, whose rank is r , if there exist two orthonormal matrices, U and W, and another diagonal matrix, D, they satisfy the following equation:

$$A_{m\times n} = U_{m\times m} D_{m\times n} W_{n\times n}{}^{T} = \sum_{i}^{r} \delta_i u_i \omega_i^{T} \qquad \text{(No)}$$

$$U^T U = E_{m\times m} \qquad W^T W = E_{n\times n}$$

Equation (2) is called the singular value decomposition of the real matrix $A_{m\times n}$.

In this equation, $U_{m\times m} = [u_1, u_2, ..., u_m]$, $D_{m\times n} = \begin{bmatrix} \Delta_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}$, $\Delta_{r\times r} = diag(\delta_1, \delta_2, ..., \delta_r)$, $W_{n\times n} = [w_1, w_2, ..., w_i]$, $r = \min(m, n)$ and $\delta_i (i = 1, 2, ..., r)$ are the singular values of the real matrix $A_{m\times n}$. $\delta_i = \sqrt{\lambda_i}$, $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_r \ge 0$,are the eigenvalues of $A^T A$. Under the restrictions of $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_r \ge 0$, the singular value of the matrix $(\delta_1, \delta_2, ..., \delta_r)$ is unique. Te singular value has the following two features: (1) the singular values of matrices have a better stability and (2) the singular values also have both proportion invariance and rotation invariance. Terefore, singular values can refect the features of eigenvectors very well. In the process of constructing the real matrix A, the time delay embedding technique is usually used to reconstruct the phase space of one-dimensional time series. However, there is no clear theoretical guidance on how to determine the embedding dimension and the delay constant (J.Gai, et al,2018).

In case of vibratory signals, the SVD of a matrix made up of vibratory measurements at different point if a structure, makes it possible under certain conditions to remove the dominant spatial eignmodes present in these data. SVD decompose the vibration signals which is collected by the acceleration sensor placed on the bearing seat, in the three above-mentioned states (normal state, slightly worn state, and severely worn state). SVD can reduce the noise components and intermittent signals. Because the singular value has high stability. Thus SVD makes it possible to “zoom in” and reveal the different characteristic frequencies of the defects which do not appear clearly during a first wavelet transform. It acts a bit like a filtering of the uniteresting components in favor of the dominant companents (X. Milisien, et al, 2006) .

II.7 Artificial Neural Network

ANN is one of the classifiers most commonly used in intelligent fault diagnosis. NN is an interconnected group of artificial neurons. These neurons use a mathematical or computational model for information processing. For its most popular form, there are three components in an ANN: input layer, hidden layer and output layer. Units in hidden layer are called hidden units, because their values are not observed. The circles labeled ‘‘+1” are intercept terms and are called bias units.ANN is an adaptive system that changes its structure based on information that flows through the network. A single neuron consists of synapses, adder and activation function. Bias is an external parameter of neural network. Model of a neuron shown in Figure, can be represented by following mathematical model: (P.K.Kankar, et al,2012) (Zurada, 1999)

$$y = f(x) = W^T x + b = \sum_{i=1}^{N} W_i x_i + b$$

where f is called the activation function, often chosen to be the sigmoid function; W are the ANN model parameters (or weights); b is a scalar. An ANN interconnects many ‘‘neurons” and the output of a neuron can be the input of another. The weights W are obtained by an iterative training procedure, based on known input–output patterns.

ANNs implement algorithms that attempt to achieve a neurological related performance, such as learning from experience, making generalizations from similar situations and judging states where poor results were achieved in the past.

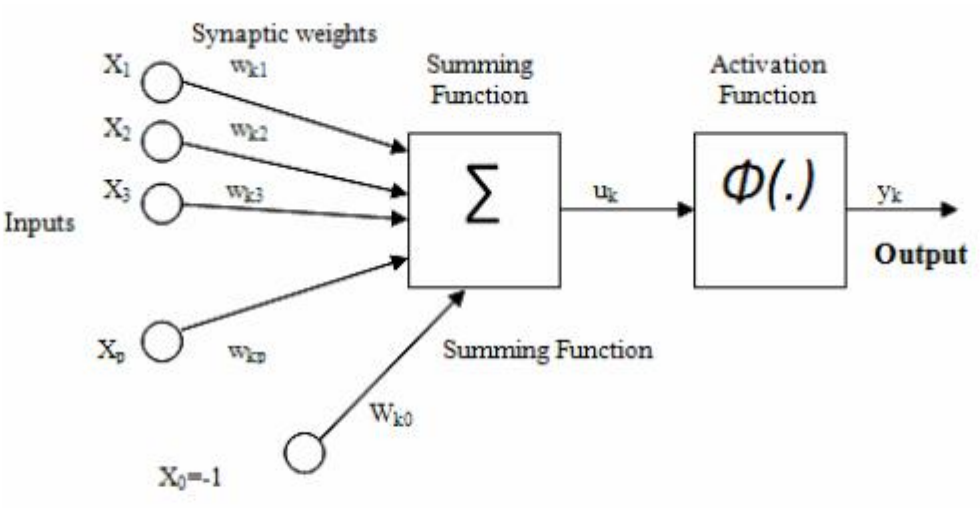

Figure 5: Model of a single non-linear neuron (P.K.Kankar, et al,2012)

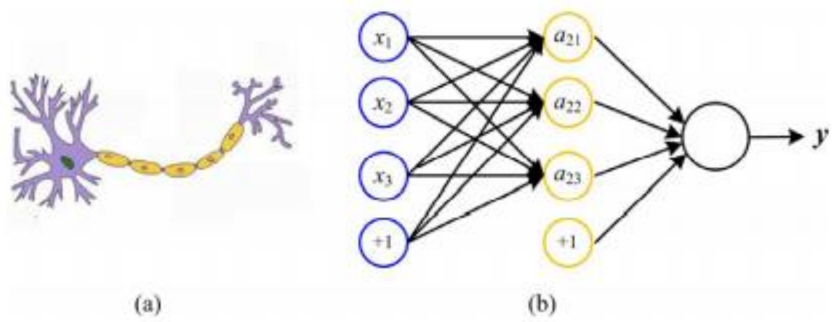

Figure 6: Human neuron and a MLP with two hidden layers(R. Liu ,et al,2018)

With neural networks, it is possible to estimate a nonlinear function without requiring a mathematical description of the functional relationship between input and output. The most commonly mentioned advantages for neural networks are their ability to build a model for any nonlinear system, the ability to learn for strongly parallel structures, and the ability to process conflicting or noisy data. Defect diagnostic domain solves many of the problems encountered when using analytical models. With neural networks, it is possible to generate the residuals and build a decision on the defect. Neural networks are used in situations where it would be possible to obtain data from measurements. The large amount of digital data in the system is also an essential condition for creating neural networks. The reliability of these networks is increased if there are not enough metrics provided by all states the operation of the system (S. Samir Hamdani,2012).

II.8 Support Vector Machine (SVM)

SVM is a supervised machine learning method based on the statistical learning theory. It is a useful method for classification and regression in small-sample cases such as fault diagnosis. Pattern recognition and classification using SVM is described in brief. A simple case of two classes is considered, which can be separated by a linear classifier. Figure 7 shows triangles and squares stand for these two classes of sample points. Hyper plane H is one of the separation planes that separate two classes. H1 and H2 (shown by dashed lines) are the planes those are parallel to H and pass through the sample points closest to H in these two classes. Margin is the distance between H1 and H2 (P.K. Kankar, et al,2011). The SVM tries to place a linear boundary between the two different classes H1 and H2, and orientate it in such a way that the margin is maximised, which results in least generalisation error. The nearest data points that used to define the margin are called support vectors (P. Konar, et al,2011).

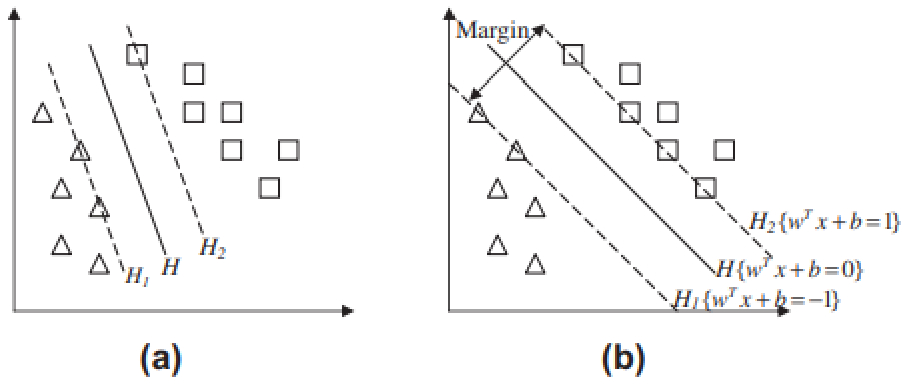

Figure 7: Hyperplane classifying two classes: (a) small margin and (b) large margin(P.K. Kankar, et al,2011)

II.9 k-Nearest Neighbour

KNN is a supervised learning algorithm and widely applied in pattern recognition. In comparison with other types of supervised learning algorithms such as neural networks, KNN is a non-parametric method because it makes no assumptions about the relationship between the outcome variable and the predictors and therefore does not involve estimation of parameters of the assumed functions nor does it need a model at the training stage. KNN assumes all observations correspond to points in the p-dimensional space. The nearest neighbors of an observation are defined in terms of the standard Euclidean distance. An observation is classified by a majority vote of its neighbors, with the observation being assigned the class most common amongst its k nearest neighbors. Suppose one has a database consisting of a total of n observations (x_i ; y_i), for i n = 1, 2,..., , where x_i could be any point in a p-dimensional. Euclidean space,$\Re^p$, denoted as $x_i=\{x_{i1}, x_{i2},\dots, x_{ip}\}$ and y_i is an outcome from m class $\omega=\{\omega_1,\omega_2,\dots,\omega_m\}$. The database is called the training set for the KNN algorithm. Given any two observations, x_i and x_j , let s(x_i , x_j) be a measure of their similarity based on the p varieties, it can be derived as (D. He, et al,2011):

$$s(x_i, x_j) = \sqrt{\sum_{k=1}^{p} (x_{ik}, x_{jk})^2}$$

To classify the response for a new observation x_u with the KNN algorithm, one first identifies k observations in the training set that are most similar to x_u . They form the set of k-nearest neighbors of x_u , denoted by N(x_u,k). These similarities can be ordered. Denote the ordered similarities with $s_{(i)}$,i.e., $s_{(1)} \geq s_{(2)} \geq\dots\geq s_{(n)}$. In other words, if $s_j=s_{(k)}$, it means x_j is the k^{th} most similar observation in the training set to x_u. The set of the k-nearest neighbors of x_u, N(x_u,k), can then be defined as all observations hose similarities to x_u are at least $s_{(k)}$,i.e., N(x_u,k) =$\{x_i:s_i\geq s_{(k)}\}$. The KNN algorithm then estimates the probability that $y_u=\omega_i$ by the average responses of these k-nearest neighbors and classifies the response to be ω_i if the estimated probability exceeds a certain threshold c:

$$\hat{P} = \frac{\sum_{x_i \in N(x_u, K)} y_i}{|N(x_u, K)|}$$

Where $|N(x_u, k)|$ is the number of items contained in the set N(x_u,k). This is usually equal to k exactly, but may exceed k depending on how ties are treated . The response is then predicted to be ω_i if $\hat{P} \geq c$, where c is a pre-specified threshold parameter. Compared with other AI algorithms, k-NN shows an advantage of simple implementation.

II.10 Naive Bayes Classifier

The NBC Theorem is a simple machine learning algorithm that is used to analyze the occurrence of an event based on the evidence or data. The algorithm is trained mainly for the classification problems in various domains, primarily used for the text classification issues like spam filtration of mail, predicting the health risk and their issues. Despite its simplicity, NB is known for its quick and effective computation of unknown class from high dimensional datasets. It can perform the classification of both supervised and unsupervised data. As its name suggests, NBC takes the assumption that presence or absence of a class is unaffected by the presence or absence of another class. The same scenario has been considered in the bearing fault cases here. As the all 3 faults of bearing, i.e., ORF, IRF, and BF have been taken as independent faults rather than dependent on the occurrence of other fault(M. K.Saini, et al,2018). In simple words, NBC theorem is based on conditional probability with the independence assumption of attributes. It is most suitable for continuous, discrete, and categorical features data sets. NBC mainly classified into three types; Multinomial Naive, Bernoulli Naive, and Gaussian Naive.

II.10.1 Multinomial or Binomial NBC Theorem

The Multinomial or Binomial Network Theorem is one of the standard classic algorithm mainly used for data classification problem. The binomial model explicitly gives the count and frequency of occurred class or event based on the given data or evidence. Thus, class regulation is possible for irregular data sets and most suited for discrete data types. When classifying the data type problem, the binomial model measures the frequency of the class. Spam detection in e-mail is the practical examples of Binomial Naive Bayes Network Theorem. This algorithm is unfit for continuous data types and random selection is not possible.

II.10.2 Bernoulli NBC Theorem

Bernoulli NBC Theorem is a special algorithm which uses binary values (0,1) for indicating the data classifications. The method also opts for discrete data and inherently it solves the data classification problem. The Boolean function indicates the presence or absence of a class in a data; 0 for absence and 1 for presence, respectively. The main principle behind the Bernoulli Theorem is conditional probabilities and firmly suitable for a small group of data. The method does computational mistake for more number of class because of its binary nature.

II.10.3 Gaussian NBC Theorem

The Gaussian NBC Theorem is a simple algorithm among the three types and mainly follows a normal distribution. The main parameters are mean and variance. In contrast to Binomial and Bernoulli Theorem, Gaussian accepts input in the form of continuous datasets. It can perform the classification for both supervised and unsupervised data. Most opted method for statistical type problem and training the data is achieved through condition independence function. Random selection is also possible with Gaussian NBC Theorem. Additionally, a kernel function is included, which performs the mapping of high input data. From the three models, Gaussian NBC Theorem is selected to accomplish the bearing failure diagnosis.

Furthermore, the algorithm is trained by importing the data from three bearing conditions and training them properly.

The derivation of the algorithm is based on the NBC rules and assumptions of conditional independence. The main equation prescribes the goal of learning is P (X|Y), where X= (X_1, X_2,…X_n). The Gaussian algorithm assumes that Xn is equal to conditional independence; conditions to be specified at the time of training. The total data of X_n is given to Y. The value of each subset P (X|Y) is calculated according to the input X and the problem of estimating the training data is neglected.

Let us calculate the Gaussian NBC model for two input models (X= X_1, X_2). In this case, the Bayes algorithm can be derived as follows:

$$P\left(X\mid Y\right) = P\left(X_1, X_2\mid Y\right)$$

In the second step, the general property of the probability is calculated.

$$P\left(X\mid Y\right) = P\left(X_1\mid X_2, Y\right)P\left(X_2\mid Y\right)$$

The last step is to calculate the definition of conditional independence.

$$P\left(X\mid Y\right) = P\left(X_1\mid Y\right)P\left(X_2\mid Y\right)$$

Generally, the Bayes network algorithm for n attributes which satisfy the conditional independence assumption is given by,

$$P\left(X_1\ldots X_n\mid Y\right)=\prod_{i=1}^{n}P\left(X_i\mid Y\right)$$

Both the Y and Xi are Boolean variables and assume of conditional independence. X takes the value of the bearing condition and Y takes the value of the amplitude of characteristic frequency components.

II.11 Fuzzy Neural Network

Fuzzy neural network is an important intelligent information processing method, which combines the advantages of fuzzy logic and neural networks well. Terefore, the fuzzy neural network algorithm not only has a strong self-learning ability to deal with data directly but also has a strong ability of structural knowledge expression. Figure 8 shows the general structure of the fuzzy neural network. In the fuzzy neural network, the function of the frst layer is to transfer the input signal to the next layer without any change. Each of its nodes corresponds to an input constant. Te number of nodes in the input layer is determined by the input signal. The second layer is the quantized input layer; its function is to fuzzify the input variables. The third layer is the hidden layer; its function is to realize the mapping between the fuzzy values of the input variables and the output variables. Te fourth layer is the quantized output layer; its output result is the fuzzifcation value.Te ffh layer is the weighted output layer; it can make sure of the clarity of the output results.

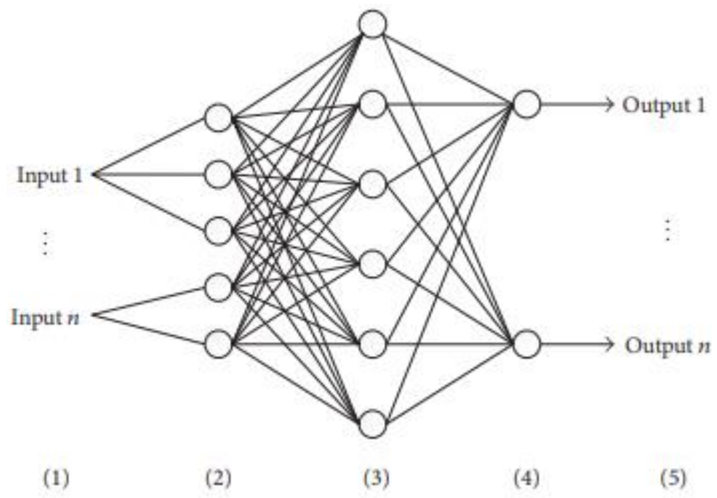

Figure 8: Structure diagram of the fuzzy neural network (J.Gai, et al,2018)

Network learning is a process of updating network connection parameters constantly, and it ultimately makes the network achieve optimal performance. The learning error is calculated according to the actual output value and the target value, and then the condition parameter is adjusted by the back-propagation error. Te specifc adjustment process is as follows.

1. Network initialization: determine the number of input nodes, hidden nodes, and output nodes of the network; initialize parameters of fuzzy neural network and fuzzy membership degree at the same time.

2. Te training of fuzzy neural network: input the sample X_n =(n=1,2,..n) and the label

Y_n =(n=1,2,..n), and then set the hyperparameter (such as the number of iterations and the network learning rate).

3. Start iteration: adjust the network connection parameters constantly until the actual output of the network is consistent with the ideal one; then, the training is over.

4. Classify the samples with the trained fuzzy neural network.

II.12 Convolutional Neural Network (CNN)

CNN is a variant of multi-layer neural network and usually consists of alternating convolutional and pooling layers . In contrast to fully-connected network, CNN presents three basic characteristics, i.e. local receptive fields, shared weights and pooling, reducing the number of trainable parameters without the loss of the expressive power. This also makes CNN easy to train with a backpropagation algorithm. CNN is a hierarchical structure, stacking multiple layers. The layer types considered in this paper are introduced in the next subsections (K. Gryllias, et al,2018).

II.12.1 Convolutional Layer

In the convolutional layer, each neuron is connected to a small region of the input neurons. The region in the input is called local receptive field. Then the convolution operation is conducted on the local receptive field to extract local features with a learned convolution kernel/weight. For each input x_i and convolution kernel k_j, the output feature map is calculated as follows:

$$y_{i,j} = f(b_j + \sum_i k_j * x_i)$$

$$f(x) = \text{ax}(0, x), \qquad x > 0$$

where, * denotes the convolution operation, k and b are the shared value of the kernel and the bias, which means that all the neurons in this layer detect the same feature, just at different locations in the input. $f(.)$ is the neural activation function, which is usually selected as the Rectified Linear Unit (ReLU) to accelerate the convergence of CNN.

II.12.2 Pooling Layer

The pooling layer is used to obtain a representation that is invariant against small translations and distortions. This is achieved by summarizing the feature responses in a region of neurons in the previous layer. For an input feature map x_i, the output feature map y_i is obtained:

$$y_i = \max_{r \times r}(x_i)$$

where r is the pooling size and the common pooling operation adopted is known as max-pooling, which simply outputs the maximum activation in the input region.

II.13 Cyclic Spectral Coherence (CSCoh)

Cyclostationary signals present some statistical properties that vary cyclically with time. Opposed to stationary signals, cyclostationary signals, though not periodic, are generated by periodic mechanism and contain extra information, which is carried by hidden periodicities. Such signals can be characterized by the second-order of cyclostationarity(Z. Chen, et al,2020). In rotating machines, the bearing defects usually generate modulated signals by the characteristic frequencies of the bearings. Such signals are usually second-order periodicity processes and are able to be separated from other interfering signals to detect and identify their hidden periodic behavior (K. Gryllias, et al,2018).

For a cyclostationary signal x(t), the second-order moment of cyclostationarity can be defined as an instantaneous AutoCorrelation Function (ACF) with a cyclic period T, which is described as:

$$R_{xx}(t,\tau) = R_{xx}(t+T,\tau) = E\{x(t+\tau/2)x(t-\tau/2)^{*}\} \quad (1)$$

where the star (*) denotes complex conjugation, s is the time-lag, and E is the statistic mean. The Fourier coefficients of the ACF correspond to the cyclic ACF and are obtained by:

$$R_{xx}(\tau,\alpha) = \int R(t,\tau)e^{-j2\pi\alpha t}dt \quad (2)$$

where a is defined as the cyclic frequency. From the Eq. (2), it can be observed that the cyclic ACF represents the Fourier coefficients of a with respect to a time-lag signal $R(t,\tau)$. The second-order statistical descriptor of cyclostationarity, called Cyclic Spectral Correlation (CSC), can be estimated by implementing the Fourier transform on the Cyclic ACF, which is given by:

$$CSC(\alpha, f) = \int R_{xx}(\tau,\alpha)e^{-j2\pi f\tau}d\tau = \int\int R(t,\tau)e^{-j2\pi(\alpha t+f\tau)}dtd\tau \quad (3)$$

The CSC can also be defined as the double Fourier transform of the signal, being a function of the spectral frequency f and the cyclic frequency α. Contrary to the classic spectral analysis, it provides an additional frequency dimension, revealing both the carriers and their modulations. Spectral frequency f is linked to the carrier component, and the cyclic frequency a is linked to its modulation. When is equal α to zero, it corresponds to the classical power spectrum. On the other hand, when α is not equal to zero, it indicates the power spectrum for that specific cyclic component. Then the CSCoh can be used to measure the degree of correlation between two spectral components estimated by:

$$CSCoh(\alpha, f) = \frac{CSC(\alpha, f)}{CSC(0, f)CSC(0, f-\alpha)} \tag{4}$$

The CSCoh can be interpreted as the CSC of a whitened signal, which tends to equalize regions with very different energy levels, magnifying weak cyclostationary signals.

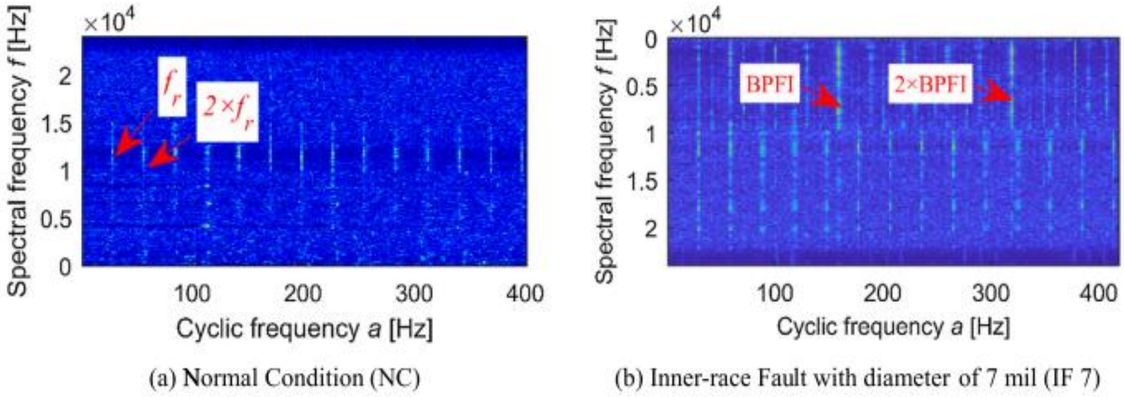

Figure 9: The 2D CSCoh maps of the different bearing health conditions(K. Gryllias, et al,2018)

III Conclusion

Fault diagnosis for rotating machinery is vital to reducing maintenance costs, operation downtime and safety hazards. In this study, a number of diagnosis techniques using for detection of bearing fault have been presented among envelope analysis, Singular value decomposition , STFT, Continuous wavelet transform, Spectral kurtosis, EMD decomposition. Moreover AI techniques have been presented for rotating machinery diagnosis include k- Nearest neighbour, Naive Bayes-based, Support vector machine and Artificial neural network as will as a deep learning methods include bConvolutional neural network and Cyclic Spectral Coherence .

REFERENCES

D. He, R. Li, and J. Zhu,” Plastic Bearing Fault Diagnosis Based on a Two-Step Data Mining Approach,”IEEE, 2011.

D. Wang ,P. W. Tse, K. L. Tsui,” An enhanced Kurtogram method for fault diagnosis of rolling element bearings,” Mechanical Systems and Signal Processing ,35 176–199, 2013.

J. Gai. Y. Hu,” Research on Fault Diagnosis Based on Singular Value Decomposition and Fuzzy Neural Network,” Shock and Vibration, Article ID 8218657, 7 pages,2018.

K. Gryllias, A. Mauricio, J. Qi, Advanced cyclostationary-based analysis for condition monitoring of complex systems, in: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, pp. 385–389.,2018

M. K.Saini, A. Aggarwal,” Detection and diagnosis of induction motor bearing faults using multiwavelet transform and naive Bayes classifier,” Int Trans Electr Energ Syst.28:e2577,2018.

N. Sawalhi, R.B. Randall, H. Endo,” The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis,” Mechanical Systems and Signal Processing 21 ,2616–2633,2007.

N.E. Huang, Z. Shen, S.R. Long, A new view of nonlinear water waves: the Hilbert spectrum, Annual Reviews of Fluid Mechanics 3 ,417–457,1999.

P. Konar, P. Chattopadhyay, Bearing fault detection of induction motor using wavelet and support vector machines (SVMs), Appl. Soft Comput. 11 (6), 4203–4211,2011.

P.K. Kankar, Satish C. Sharma, S.P. Harsha,” Fault diagnosis of ball bearings using machine learning methods,” Expert Systems with Applications 38, 1876–1886,2011.

P.K.Kankar, S.C. Sharma, S. P.Harsha,” Vibration-based fault diagnosis of a rotor bearing system using artificial neural network and support vector machine,” Int. J. Modelling, Identification and Control, Vol. 15, No. 3, 2012

R. Liu , B. Yang ,E. Zio , X. Chen,” Artificial intelligence for fault diagnosis of rotating machinery: A review,” Mechanical Systems and Signal Processing 108,pp. 33–47,2018

S. Prabhakar, A.R. Mohanty, A.S Sekhar,”Application of discrete wavelet transform for detection of ball bearing race faults ,”Tribology International 35,793-800,2002.

S.Karmakar, S.Chattopadhyay, M. Mitra, S.Sengupta,” Induction Motor Fault Diagnosis: Approach through Current Signature Analysis,” Springer Science,2016.

X. MILISEN, V. LENAERTS,” Comparaison de différentes méthodes de détection de défauts dans les roulements,” Revue Scientifique des ISILF n°20, 2006.

Y. Yu, Y.Dejie, Ch.Junsheng,” A roller bearing fault diagnosis method based on EMD energy entropy and ANN,” Journal of Sound and Vibration 294, 269–277,2006.

Z. Chen , A. Mauricio , W. Li , K. Gryllias," A deep learning method for bearing fault diagnosis based on Cyclic Spectral Coherence and Convolutional Neural Networks," Mechanical Systems and Signal Processing (140) 106683,2020

S. Mohanty, K. K Gupta, K.S. Raju, A. Singh, S.Snigdha "Vibro Acoustic Signal Analysis in Fault Finding of Bearing Using Empirical Mode Decomposition," International Conference on Advanced Electronic Systems,2013.

X. Guo, L. Chen, C. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement (2016) 490–502,2016.

R. Rubini, U. Meneghetti," Application of the envelpe and wavelet transform abnalyses for the diagnosis of incipient faults in ball bearings," Mechanical Systems and Signal Processing 15(2), 287-302,2001.