Vibration Signal Based Multi-fault Diagnosis of Gears using Roughset Integrated PCA and Neural Networks

S.Devendiran*1, K.Manivannan2
1,2School of Mechanical and Building Sciences, VIT University, Vellore, Tamil Nadu, India
devendiran@vit.ac.in
manivannan.k@vit.ac.in

Abstract--- Fault diagnosis in gears has been the subject of intensive research as gears are critical element in a variety of rotating machinery applications and an unexpected failure of the gear may cause substantial economic losses. Vibration analysis has been used as a predictive maintenance procedure and as a support for machinery maintenance decisions. In this present work monitoring of gear health using the vibration signals as well as usage of advance signal processing methods such as wavelet transform is implemented to extract the features which reveals the characteristics of the actual conditions of the gear. By measuring and analysing the signals, it is possible to determine both the nature and severity of the defect, and hence predict the machine’s failure. To reduce the dimensionality of the extracted statistical features, roughset integrated principle component analysis (RSPCA) is used after feature extraction process. Back-propagation neural network (BPNN) and probabilistic neural network (PNN) were deployed for diagnostic analysis of the signal. To validate the proposed methodology, four kinds of running states are simulated with artificially created faults in gear that is accommodated in the test rig. By comparing classification accuracy along with computation timing the best scheme is selected for diagnostic prediction.

Index Term-- Test rig, fault diagnosis, wavelet, roughset, principle component analysis (RSPCA), back-propagation neural network (BPNN) and probabilistic neural network (PNN).

I. INTRODUCTION

The occurrence of failures in the machinery can be costly and even catastrophic. In order to avoid them, there needs to be a system or a methodology by which, analysis of the behaviour of the machine is done and thereby provide an alarm and instructions for preventive maintenance. Analysing the behaviour of the machines has become possible by means of advanced sensors, data collection systems, data storage/transfer capabilities and data analytic tools developed for such purpose. Fault prediction is one of the important undertakings during the maintenance work; especially in a gearbox early gear fault prediction is help prevent damage of the entire system. Vibration signal based gear fault diagnosing method is prominently used over the past two decades [1]. In Particular, time domain signal analysis as well frequency domain analysis by Fast Fourier transform (FFT) is used for the detection of gear failures by many of the researchers [2]. When FFT is considered for non-stationary signals the frequency component will depends on the time and it will vary therefore the FFT is not suitable for fault diagnosis in this aspect [3]. For analysing the vibration signals and extraction of faulty features of rotating machines is done by using the latest signal processing techniques based on time and frequency domain technique which provide both the time and frequency information of the vibration signal such as Wavelet Transform (WT) [4]. The Wavelet Transform (WT) provides powerful multi-resolution analysis in both time and frequency domains and thereby becomes a favoured tool to extract the transient signal components/ characteristics features of the non-stationary vibrations signals produced by the faults. On the basis of extracted wavelet coefficients the analysis is done and the results are affected by choice of wavelet basis, thus by this reason choice of wavelet can be affected and leads to assumption of signal. Hilbert Huang Transform is a method [5] which is based on the local characteristic time scales of a signal and could decompose the complicated signal into a number of intrinsic mode functions. EMD is another self-adaptive method through which the IMF (intrinsic mode function) is obtained by reducing the raw data. IMFs gives the natural oscillatory modes which is entrenched in the raw signal and hence it is widely used in fault detection in machines [6]. Some of the researchers [7] made attempt on detection of gear condition using the ANN and the SVM with the Genetic algorithm (GA) based attribute collection from vibration data [8]. After signal processing, statistical characteristics such as mean, peak to peak, standard deviation, skewness, kurtosis, and crest factor were calculated for different states of fault conditions. The vast number of features within variety domains poses challenges to data mining. In order to achieve successful data mining, feature selection is an essential process. Many researchers utilized GA for reduce the dimensionality of the features by selecting an optimized features [9-11]. GA also used along with ANFIS [12] in rotating machinery classification process. The feature selection such as principal component analysis (PCA) [13] and J48 algorithm [14] are widely used to decrease dimensions of features and proved for its meaningful diagnostics of conditions of faults in gear components.
In Section 1 we briefly reviewed the works which are strictly connected to the subject of this paper. In Section 2 Experimental setup and experimental procedure is presented. In Section 3 Theoretical background of wavelet and application of the most famous signal processing technique based on sub-band coding and known for its fast computation among the wavelet family called discrete wavelet decomposition is used for extracting wavelet coefficients and also statistical feature extraction process for different levels of decomposition were presented. In Section 4 Theoretical background of Roughset integrated PCA and its mathematical process behind the feature selection were presented. In Section 5 Theory behind the neural network BPNN and PNN is given. In Section 6 overall implementation process such as feature extraction, feature selection, classification, its result analysis and discussion were presented and Last section contains conclusions. The methodology of the present work is illustrated in the Fig 2.

II. EXPERIMENTAL SETUP

Experimental setup consist of a three phase 0.5hp AC motor, a variable frequency drive (VFD) for regulating the speed of the motor, tri axial accelerometer as a sensor and data acquisition system integrated with computer as in Fig 1. The motor is attached to a shaft and connected to the gearbox with the belt drive. The speed of motor is maintained at constant speed (1700 rpm) and as well the load, to acquire the vibration signal from the gearbox. The load is controlled by brake drum dynamometer. The tri axial accelerometer is placed on the top of the gearbox and connected to the DAS (data acquisition system) and DAS is connected to the computer where the vibration signal data sets are stored and analyse. Gear box consists of spur gear and pinion with the specification of 36 and 24 teeth respectively, module 3mm and a pressure angle of 20°. Different gear condition such as normal, fault1 (frosting), fault2 (pitting), fault3 (crack) is artificially created. The vibration data values for different condition of gear condition is acquired and stored for the further process.

III. WAVELET TRANSFORM AND FEATURE EXTRACTION

The next step after experimentation is signal processing and it is continued with feature extraction. The aim of signal processing is to find the health condition of the rotating component using vibration signal. Recently, wavelet transform (WT), which is capable of providing both the time-domain and frequency-domain information simultaneously, has been successfully used in non-stationary vibration signal processing and fault diagnosis [15]. In this present work wavelet transform is used in feature extraction. The two type of wavelet Continuous wavelet transform (CWT) and Discrete wavelet transform (DWT) are discussed below.

A. Continuous Wavelet Transform (CWT)

A series of oscillating functions with different frequencies such as window function \( \psi_{a,\beta}(t) \) for scanning, \( x(t) \) for translating the signal are used in wavelet transform or linear transform, where \( a \) is the oscillating frequency changing parameter and \( \beta \) is the translation parameter.
A high frequency resolution is obtained when the frequency is low at low time resolution and a low frequency resolution is obtained when the frequency is at low time resolution.

The function with the parameter \( \alpha \) and \( \beta \) with mother wavelet is given as:

\[
\Psi_{\alpha, \beta}(t) = \frac{1}{\sqrt{\alpha}} \Psi\left(\frac{t-\beta}{\alpha}\right)
\]  

(1)

If all signals \( x(t) \) satisfy the condition then,

\[
\int_{-\infty}^{\infty} |x(t)|^2 dt < \infty
\]

(2)

that gives \( x(t) \) decays to zero.

The wavelet transform, \( CWT(\alpha, \beta) \), of a time signal \( x(t) \) can be given as

\[
CWT(\alpha, \beta) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty} x(t) \Psi^*(\frac{t-\beta}{\alpha}) dt
\]

(3)

Where an analysing wavelet is \( \Psi(t) \) and \( \Psi^*(t) \) will be the complex conjugate of \( \Psi(t) \). For analyzing the different wavelet, real and complex valued function can be used. As given in Equation (1), the transform signal \( CWT(\alpha, \beta) \) is defined on plane \( \alpha, \beta \), where \( \alpha \) and \( \beta \) are used to change the frequency and the time location of the wavelet. Whenever high frequency resolution is required, the decrement of \( x \) will construct a high-frequency wavelet and vice versa is possible. In other side as \( y \) increases, the wavelet traverse the length of the input signal, and increases or decreases in response to changes in the local time and frequency content of the signals.

B. Discrete Wavelet Transform (DWT)

A lot of redundant data is created by calculating the wavelet coefficient at possible scale. Thus the analysis can be made sufficient and accurate by giving the limit of choice of \( \alpha \) and \( \beta \) to discrete numbers.

By selecting the fixed values \( \alpha = \alpha_0^j \) and \( \beta = k\beta_0\alpha_0^j \), \( j, k = 0, \pm 1, \pm 2, ..., \) we get the DWT

\[
DWT(i,j) = \alpha_0^{-\frac{j}{2}} \int_{-\infty}^{\infty} x(t) \Psi^*\left(\frac{\alpha_0^{-\frac{j}{2}} t - k\beta_0}{\alpha_0^j}\right) dt
\]

(4)

If \( \alpha \) and \( \beta \) are replaced by \( 2^j \) and \( 2^j k \), then the DWT is derived as

\[
DWT(i,j) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} x(t) \Psi^*\left(\frac{t-2^jk}{2^j}\right) dt
\]

(5)

Thus the high and low frequency signal is obtained from the original signals by decomposition process. The daubechies wavelet transform is used in this process because of its less process consuming time. As well it is in the family of orthogonal wavelet.

This is a type of multi resolution analysis and mostly used for solving a wide range of problems with large dataset. The scaling function of daubechies wavelet is given as db1-db15. It is noted that the highest level of efficiency is taken for the analysis by the Daubechies wavelet.

![Fig. 3(a). Decomposed signal of normal gear using db15scale](image1)

![Fig. 3(b). Decomposed signal of frosting gear using db15scale](image2)
The gear signals are extracted in the sampling rate of 12800hz (6400 data points per second). For 20 seconds 128000 data points were collected through accelerometer for each condition of gear. Then the raw signals are divided into approximately around 21 sample datasets and each sample contain 6000 data points were given as input to the wavelet transform and decomposed using daubechies wavelet for 10 level and 15 scale as show in Figs 3(a), 3(b), 3(c), 3(d) respectively for four classes of gears. By trial and error the scales of wavelet is fixed based of the result accuracy. From the decomposed data the statistical time domain features are obtained using Table 1. These features are useful in fault prediction and are further processed and normalized. Sample feature (STD) values are tabulated for the decomposed signal is shown in Table 2.

**IV. PRINCIPAL COMPONENT ANALYSIS (PCA)**

In order to reduce the number of features and dimensions of the features without affecting the usefulness of the features, feature reduction method is used.

The given set of n dimension feature vectors \( x_t(t=1,2,3,\ldots,m) \) generally \( n < m \)

\[
\mu = \frac{1}{m} \sum_{t=1}^{m} x_t
\]  

Then by using the formula the covariance matrix of feature vector is found

\[
C = \frac{1}{m} \sum_{t=1}^{m} (x_t - \mu)(x_t - \mu)^t
\]  

By solving the eigen value problem of covariance matrix \( C \), The principal components (PCs) is obtained

\[
Cv_i = \lambda_i v_i
\]

Where \( \lambda \) (i = 1, 2, 3, ..., n) are the eigen values and they are sorted in descending order, \( v_i \) (i = 1, 2, ..., n) are the corresponding eigenvectors. the raw feature vectors with low-dimensional ones, what needs to be done is to compute the first k eigenvectors \( k \leq n \) which correspond to the biggest k eigen values. In order to select the number k, a threshold \( \theta \) is introduced to denote the biggest k eigenvectors.

\[
\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \geq \theta
\]

Given the precision parameter \( \theta \), the number of eigenvectors k can be decided.

Let,

\[
v = [v_1, v_2, \ldots, v_k] = \text{diagonal}[\lambda_1, \lambda_2, \ldots, \lambda_k]
\]
After the matrix V is decided, the low-dimensional feature vectors, named principle component (PC), of raw ones are obtained as follows,

\[ P = V^T x_t \]  

The Principal Component’s (PCs) of PCA have three properties[8]. 1) They are uncorrelated; 2) They have sequentially maximum variances; 3) The mean-squared approximation error in the representation of the original vectors by the first several PCs is minimal.

**A. Rough Set Theory**

RST (rough set theory) have been used in many domains for research purposes in the past decade. Particularly in reduction of attributes (features) in a dataset the RST is used [16]. The roughset consist of inner boundary and the outer boundary so the refined values are obtained from the region in between the two boundaries (Fig 4).

![Fig. 4. Schematic representation of roughset](image)

### Table II

<table>
<thead>
<tr>
<th>STATES</th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
<th>d5</th>
<th>d6</th>
<th>d7</th>
<th>d8</th>
<th>d9</th>
<th>d10</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORMAL</td>
<td>0.797528</td>
<td>0.033356</td>
<td>0.933221</td>
<td>0.956247</td>
<td>0.888151</td>
<td>0.911582</td>
<td>0.871224</td>
<td>0.962043</td>
<td>0.114847</td>
<td>0.261767</td>
</tr>
<tr>
<td>NORMAL</td>
<td>0.797749</td>
<td>0.019003</td>
<td>0.995923</td>
<td>0.989622</td>
<td>0.913572</td>
<td>1</td>
<td>0.975675</td>
<td>1</td>
<td>0.027988</td>
<td>0.301418</td>
</tr>
<tr>
<td>NORMAL</td>
<td>0.992271</td>
<td>0.234528</td>
<td>0.017552</td>
<td>0.17678</td>
<td>0.055278</td>
<td>0.131438</td>
<td>0.066551</td>
<td>0.059223</td>
<td>0.946283</td>
<td>0.431604</td>
</tr>
<tr>
<td>NORMAL</td>
<td>0.074115</td>
<td>0.946283</td>
<td>0.084145</td>
<td>0.053661</td>
<td>0.127252</td>
<td>0.017336</td>
<td>0.007075</td>
<td>0.000769</td>
<td>0.929184</td>
<td>0.340936</td>
</tr>
<tr>
<td>FAULT 3</td>
<td>0.682706</td>
<td>0.104847</td>
<td>0.944207</td>
<td>1</td>
<td>0.801908</td>
<td>0.922111</td>
<td>0.937042</td>
<td>0.968459</td>
<td>0.003873</td>
<td>0.019356</td>
</tr>
<tr>
<td>FAULT 3</td>
<td>0.811757</td>
<td>0.007988</td>
<td>0.915009</td>
<td>0.999804</td>
<td>0.776428</td>
<td>0.831351</td>
<td>0.950092</td>
<td>0.99864</td>
<td>0.01304</td>
<td>0.801083</td>
</tr>
<tr>
<td>FAULT 3</td>
<td>0.793278</td>
<td>0.439604</td>
<td>0.029546</td>
<td>0.003873</td>
<td>0.067133</td>
<td>0.001689</td>
<td>0.00187</td>
<td>0.013233</td>
<td>1</td>
<td>0.1579151</td>
</tr>
<tr>
<td>FAULT 3</td>
<td>0.930424</td>
<td>0.340936</td>
<td>0.032619</td>
<td>0.00304</td>
<td>0.04033</td>
<td>0.001622</td>
<td>0.017374</td>
<td>0.013405</td>
<td>0.131438</td>
<td>0.155791</td>
</tr>
<tr>
<td>FAULT 3</td>
<td>1</td>
<td>0.287196</td>
<td>0.054465</td>
<td>0</td>
<td>0.02934</td>
<td>0.002859</td>
<td>0.017578</td>
<td>0.012292</td>
<td>0.093215</td>
<td>0.172861</td>
</tr>
</tbody>
</table>

The unwanted minimum information attribute are removed. The tuple is the name given for the information (feature).

\[ T = (U, A) \]  

(12)

Where U and A are two finite non empty sets, U – universe of primitive object, A-set of attributes. The domain of ‘a’ is represented as the feature \( a \in A \). The partition of set A is made into two subsets namely C & D which is known as decision attributes [17]. Let \( P \subseteq A \) be a subset of attributes. The indiscernibility relation, denoted by \( IND(P) \), is an equivalence relation defined as:

\[ IND(P) = \{(x, y) \in U \times U : \forall a \in P, a(x) = a(y)\} \]  

(13)

Where \( a(x) \) represents the value of feature of a object \( x \). P is indiscernible to \( x \) and \( y \) if \( (x, y) \in IND(P) \). The family of all classes of IND (P) (Partition of U determined by P) is denoted by U/IND (P). Each element in U/IND (P) is a set of indiscernible objects with respect to P.
The classes U/IND(C) and U/IND (D) are called decision class and can be calculated by using,

\[ \frac{U}{\text{IND}(P)} = \otimes \left\{ a \in P : \frac{U}{\text{IND}(a)} \right\} \]  \hspace{1cm} (14)

Where,

\[ A \otimes B = \{ X \cap Y : \forall X \in A, \forall Y \in B, X \cap Y \neq \emptyset \} \]  \hspace{1cm} (15)

The rough set is a classical set which consist of lower and upper approximation concept, where the lower approximation consist of all the needed element in the limit and upper approximation has those belong to the concept.

\[ PX = \{ x | x \in U | \exists X \subseteq U \} \]  \hspace{1cm} (16)

\[ PX = \{ x | x \in U | X \neq \emptyset \} \]  \hspace{1cm} (17)

If P and Q is the relation of U then, the negative and positive region is termed as

\[ POS_P (Q) = UX \in U / QPX \]  \hspace{1cm} (18)

The reduced attribute is same as that of decision feature as That of original. \( R_{\min} \) is the conditional attribute set C so that

\[ \gamma (D) = \gamma (C, D) \]  \hspace{1cm} (21)

\[ R = \{ X : X \subseteq C, \gamma (X, D) = \gamma (C, D) \} \]  \hspace{1cm} (22)

\[ R_{\min} = \{ X : X \in R, Y \in R, |X| |Y| \} \]  \hspace{1cm} (23)

**B. Roughset integrated PCA**

The feature reduction of a high dimensional data into a low dimensional with minimum loss of information is processed in principle component analysis which is a unsupervised linear feature reduction method. Mainly the maximum variance in the data is found out through PCA. The PCA can be combined with roughest for getting discriminative features which is best for classification. The rough set is that which further reduces attributes in dataset according to the upper and lower approximation for reduction. The Fig 5 shows the algorithm of roughset PCA.

**Algorithm-1 : PCA and rough set based quick reduct**

Input: \( X_t (t=1,2,3,\ldots, m) \)
Output: reduced feature \( X \)

**Step-1:** \( X_t \) create \( N \times d \) data matrix, with one row vector \( x_n \) per data point. (Eqn-)

**Step-2:** \( X_t \) subtract mean \( x \) from each row vector \( x_n \) in \( X_t \)

**Step-3:** \( \Sigma \) covariance matrix of \( X_t \)

**Step-4:** Find eigenvectors and eigen values of \( \Sigma \).

**Step-5:** PC’s ← the M eigenvectors with largest eigen values.

**Step-6:** \( P \) Output PCs.

**Step 7:**

// \( S \) - set of all independent feature of \( P \); \( T \) is the dependent feature of \( P \)

Evaluate the dependency factor of \( (PS) \) with \( (PT) \), i.e., \( \gamma_s (T) \)

Evaluate the dependency factor of each independent attributes \( (S_i) \) with \( T \), i.e., \( \gamma_{(S_i)} (T) \)

Compare \( \gamma_{(S_i)} (T) \) with \( \gamma_{(S)} (T) \)

If

\[ \gamma_{(S_i)} (T) = \gamma_{S} \]
Then

put \( (S_i) \) in the null reduced set \( X \)

\[ \gamma_{(S_i)} (T) \neq \gamma_{S} \]
Then

Put \( (S_i) \) in giving highest dependency \( (S_H) \)

Evaluate the dependency factor of \( T \) with all possible combination i.e., \( S_H (i-H) \)

Compare with \( \gamma_{(S)} (T) \)

If

\[ S_H (i-H) = \gamma_{(S)} (T) \]
then

put \( S_H (i-H) \) in the null reduced set \( X \)

Else

Repeat steps Until \( \gamma_{(S)} (T) = \gamma_{(X)} (T) \)

Where \( S \) is the set of independent attributes and \( X \) is the subset of independent attributes i.e., reduced set

. Fig. 5. Pseudo code for roughset PCA
V. BACK PROPAGATION NEURAL NETWORK (BPNN)

An artificial neural network is that which resembles the model of the human brain. It is a type of information processing system like biological neural system. ANN consists of three layers as inner layer, outer layer and hidden layer. (Fig 6). Based on the dimension of the training sample the hidden layers can be one or more in number [18] BPNN (Back propagation neural network) is a associate type in ANN which contains the training and testing process. The particular minimum maximum normalization range is scaled to increase accuracy and efficiency to perform the linear transformation [19]. The Three steps are involved in training process 1) input feed forward 2) calculating the error in back propagation mode and arithmetic operation 3) alteration of weight.[20].

i. The weights are set with small random values.

ii. Input signal \( x_i \) is given to input unit \( (X_i, i=1...n) \) and propagated to the units of hidden layers.

iii. Summation of each hidden unit \( (h_j, j=1...p) \) of its weighted input signals,

\[
h_{inj} = V_{oj} + \sum_{i=1}^{n} x_i V_{ij}
\]

iv. Comparing its output signal applying its activation function and it sends the signals to output unit layers \( h_j = f(h_{inj}) \)

v. Summation of weighted input signals to each output unit \( (O_k, k=1...m) \), and applies its activation function to compute its output signal.

\[
o_{in} = w_{ok} + \sum_{j=1}^{n} h_j w_{jk}
\]

\[
o_k = f(o_{in})
\]

vi. Computation of its error information term, as each output unit \( (O_k, k=1...m) \) receives a target pattern corresponding to the input training pattern

\[
\delta_k = (t_k - o_k) f'(o_{in})
\]

\[
\Delta w_{jk} = \alpha \delta_k h_j
\]

\[
\Delta w_{ok} = \alpha \delta_k
\]

vii. Each hidden unit \( (Z_j, j=1...p) \) sums its \( \delta \) inputs,

\[
\delta_{inj} = \sum_{k=1}^{m} \delta_k w_{jk}
\]

viii. To calculate its error information the above equation is multiplied with activation function derivative.

\[
\delta_i = \delta_{inj} f'(h_{inj})
\]

ix. Calculates its weight correction term, and calculation of its bias correction term,

\[
\Delta v_i = \alpha \delta i
\]

\[
\Delta v_{oj} = \alpha \delta
\]

x. Weight Adjustment by updating each output unit \( (Y_k, k=1...m) \) in its bias and weights \( (j=0...p) \):

\[
w_{jk}(new) = w_{jk}(old) + \Delta w_{jk}
\]

xi. Updating bias and weights \( (i=0...n) \) in each hidden unit \( (Z_j, j=1...p) \) :

\[
v = v_{o}(old) + \Delta v_{oj}
\]

xii. Stop the test condition

Repeat the process if test condition fails

---

A. Probabilistic Neural Networks (PNN)

The architecture of PNN is feed forward in nature which is similar to back propagation, but differs in the way that learning occurs when compare to BPNN. PNN is supervised learning algorithm but includes no weights in its hidden layer. Instead each hidden node represents an example vector, with the example acting as the weights to that hidden node. These are not adjusted at all. Fig. 7 illustrates a schematic representation of PNN.
Basically, PNN consists of an input layer, which represents the input pattern or feature vector. The input layer is fully interconnected with the hidden layer, which consists of the example vectors (the training set for the PNN). The actual example vector serves as the weights as applied to the input layer. Finally, an output layer represents each of the possible classes for which the input data can be classified. However, the hidden layer is not fully interconnected to the output layer. The example nodes for a given class connect only to that class’s output node and none other.

Probabilistic Neural Networks (PNN) is used for the fault identification and classification of the problem [21]. This works on the basis of Bayer’s Rule (Eqn.24) by giving the probability of fault conditions.

\[
P(H_i | E) = \frac{\left( \frac{P(E | H_i)}{P(E | H_0)} \right) P(H_i)}{\sum_{h=1}^{K} \left( \frac{P(E | H_h)}{P(E | H_0)} \right) P(H_h)}
\]

where,

\[
P(H_i | E) = \text{probability of fault } H_i \text{ in symptom } E.
\]

\(P(H_i)\) = a-priori probability of fault \(H_i\) in absence of specified symptom.

\(K\) = the number of faults.

The input layer in probabilistic neural network consists of two hidden layer and outer layer. Thus the difference between the PNN to Back propagation network is it can be created after the training data sets are passed in it. From the estimation of probability density functions based on training patterns the activation function is derived.

VI. EXPERIMENTAL PROCESS, ANALYSIS AND DIAGNOSIS OF RESULTS

A. Fault identification based on RS integrated PCA, BPNN & PNN

The gear signals are extracted in the sampling rate of 12800 Hz (6400 data points per second). For approximately 5 seconds for each gear conditions, 32000 data points are collected through accelerometer. Gear fault diagnostics using discrete wavelet, statistical features are extracted in both time and frequency domain which is the basis for input to BPNN and PNN.

Input features are normalized to avoid the change in signal-magnitude that occurs due to variations in speed of rotation or some other changes.

For feature selection, rough set PCA is implemented using Matlab. The Selected features are the basis for BPNN and PNN. The experiment was carried out with and without feature selection schemes.

The training speed of ANN, requiring epochs less than 50, is enhanced using the relevant features of the signals characterising the gear conditions.

<table>
<thead>
<tr>
<th>Feature</th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
<th>d4</th>
<th>d5</th>
<th>d6</th>
<th>d7</th>
<th>d8</th>
<th>d</th>
<th>d10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS</td>
<td>F1</td>
<td>F2</td>
<td>F3</td>
<td>F4</td>
<td>F5</td>
<td>F6</td>
<td>F7</td>
<td>F8</td>
<td>F9</td>
<td>F10</td>
</tr>
<tr>
<td>KUR</td>
<td>F11</td>
<td>F12</td>
<td>F13</td>
<td>F14</td>
<td>F15</td>
<td>F16</td>
<td>F17</td>
<td>F18</td>
<td>F19</td>
<td>F20</td>
</tr>
<tr>
<td>SKW</td>
<td>F21</td>
<td>F22</td>
<td>F23</td>
<td>F24</td>
<td>F25</td>
<td>F26</td>
<td>F27</td>
<td>F28</td>
<td>F29</td>
<td>F30</td>
</tr>
<tr>
<td>VAR</td>
<td>F31</td>
<td>F32</td>
<td>F33</td>
<td>F34</td>
<td>F35</td>
<td>F36</td>
<td>F37</td>
<td>F38</td>
<td>F39</td>
<td>F40</td>
</tr>
<tr>
<td>STD</td>
<td>F41</td>
<td>F42</td>
<td>F43</td>
<td>F44</td>
<td>F45</td>
<td>F46</td>
<td>F47</td>
<td>F48</td>
<td>F49</td>
<td>F50</td>
</tr>
</tbody>
</table>

After obtaining the gear signals for four conditions using wavelet transform, statistical features such as Root mean square(RMS), kurtosis(KUR), skewness (SKW), variance(VAR), standard deviation(STD) are extracted (shown in Table 3) which acts as an input to BPNN and PNN. The schematic representation of neural network as shown in Fig. 8, comprise of input layer, two hidden layers and the output layer. The number of neurons in the first
hidden layer contains 50 which denote the normalized input features. The second layer neurons varied from 1 to 30. The number of output nodes contains four neurons and the target value of four output nodes has binary values of four conditions of gear. MATLAB neural network toolbox is used to train the back propagation neural network with Levenberg–Marquardt algorithm.

![Image](image.png)

**Fig. 8.** Variation of the MSE of the trained data as a function of Epochs (50-30-4-4)

<table>
<thead>
<tr>
<th>Run</th>
<th>Topology</th>
<th>Regression analysis MSE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training</td>
<td>Validation</td>
</tr>
<tr>
<td>1</td>
<td>50-1-4-4</td>
<td>0.9535</td>
<td>0.99265</td>
</tr>
<tr>
<td>2</td>
<td>50-2-4-4</td>
<td>0.96352</td>
<td>0.95150</td>
</tr>
<tr>
<td>3</td>
<td>50-3-4-4</td>
<td>0.98128</td>
<td>0.98810</td>
</tr>
<tr>
<td>4</td>
<td>50-4-4-4</td>
<td>0.98658</td>
<td>0.98352</td>
</tr>
<tr>
<td>5</td>
<td>50-5-4-4</td>
<td>0.99538</td>
<td>0.9945</td>
</tr>
<tr>
<td>6</td>
<td>50-6-4-4</td>
<td>0.98529</td>
<td>0.97965</td>
</tr>
<tr>
<td>7</td>
<td>50-7-4-4</td>
<td>0.99287</td>
<td>0.98965</td>
</tr>
<tr>
<td>8</td>
<td>50-8-4-4</td>
<td>0.98504</td>
<td>0.98830</td>
</tr>
<tr>
<td>9</td>
<td>50-9-4-4</td>
<td>0.99681</td>
<td>0.96976</td>
</tr>
<tr>
<td>10</td>
<td>50-10-4-4</td>
<td>0.99745</td>
<td>0.94358</td>
</tr>
<tr>
<td>11</td>
<td>50-11-4-4</td>
<td>0.99615</td>
<td>0.97765</td>
</tr>
<tr>
<td>12</td>
<td>50-12-4-4</td>
<td>0.99815</td>
<td>0.96625</td>
</tr>
<tr>
<td>13</td>
<td>50-13-4-4</td>
<td>0.99548</td>
<td>0.92745</td>
</tr>
<tr>
<td>14</td>
<td>50-14-4-4</td>
<td>0.99880</td>
<td>0.96995</td>
</tr>
<tr>
<td>15</td>
<td>50-15-4-4</td>
<td>0.9956</td>
<td>0.94935</td>
</tr>
<tr>
<td>16</td>
<td>50-16-4-4</td>
<td>0.99662</td>
<td>0.96624</td>
</tr>
<tr>
<td>17</td>
<td>50-17-4-4</td>
<td>0.99315</td>
<td>0.92754</td>
</tr>
<tr>
<td>18</td>
<td>50-18-4-4</td>
<td>0.99475</td>
<td>0.93935</td>
</tr>
<tr>
<td>19</td>
<td>50-19-4-4</td>
<td>0.99937</td>
<td>0.94475</td>
</tr>
<tr>
<td>20</td>
<td>50-20-4-4</td>
<td>0.99532</td>
<td>0.97985</td>
</tr>
<tr>
<td>21</td>
<td>50-21-4-4</td>
<td>0.98545</td>
<td>0.95874</td>
</tr>
<tr>
<td>22</td>
<td>50-22-4-4</td>
<td>0.96523</td>
<td>0.96542</td>
</tr>
<tr>
<td>23</td>
<td>50-23-4-4</td>
<td>0.99214</td>
<td>0.97635</td>
</tr>
<tr>
<td>24</td>
<td>50-24-4-4</td>
<td>0.98874</td>
<td>0.97542</td>
</tr>
<tr>
<td>25</td>
<td>50-25-4-4</td>
<td>0.99615</td>
<td>0.97765</td>
</tr>
<tr>
<td>26</td>
<td>50-26-4-4</td>
<td>0.96878</td>
<td>0.97231</td>
</tr>
<tr>
<td>27</td>
<td>50-27-4-4</td>
<td>0.94521</td>
<td>0.93245</td>
</tr>
<tr>
<td><strong>28</strong></td>
<td><strong>50-28-4-4</strong></td>
<td><strong>0.97258</strong></td>
<td><strong>0.98245</strong></td>
</tr>
<tr>
<td>29</td>
<td>50-29-4-4</td>
<td>0.90254</td>
<td>0.98054</td>
</tr>
<tr>
<td>30</td>
<td>50-30-4-4</td>
<td>0.99601</td>
<td>0.95186</td>
</tr>
</tbody>
</table>
The initial weights and biases of the network are controlled by the BPNN program. For experimentation, five statistical feature values are evaluated. For training, 75% of each feature set contains four different conditions of gears are used. Normalised features are used as an input to avoid the signals change in magnitude due to variations in speed of rotation or some other changes.

The remaining 25% of input features were used for testing. By trial and error, during training stage the configurations are modified by changing the number of hidden layer neurons and obtain the results. The target value of the first output node for the normal gear condition was set 1000 which indicate normal gear, 2nd neuron set to 0100 which indicate spalling fault gear, 3rd neuron set as 0010 which indicate pitting gear and 4th neuron set to 0001 indicate cracked gear.

A mean square error of $10^{-6}$ a minimum gradient of $10^{-10}$ and maximum iteration number (epoch) of 100 are used. The training process would stop if any of these conditions were met. The initial weights and biases of the network were generated automatically by the program. The results of training and testing the diagnostic capability of the BPNN shows that for four feature sets (RMS, KUR, SKW & VAR), success rate is 100% in the training whereas in testing, success is 91.25%, 93.85%, 95.50% & 98.75% respectively. Among them the STD based input features shows better results.

From Table 4, Network topology 50-28-4-4 is selected based on the overall regression analysis as well the least MSE and hence, this architecture is finally considered. Similar inputs are used for PNN and results were compared as shown in Table 5. The classification accuracy based on prediction is shown in Fig 9 for PNN,BPNN.
VII. CONCLUSION

An automatic fault classification scheme based on neural network, DWT wavelet and roughset PCA is presented. The results of classification algorithms BPNN and PNN are discussed along with feature reduction and without feature reduction process. From the results, it is seen that BPNN with PCA roughset reduction (BPNN-S2) gives more classification accuracy of 97.25% in comparison with other schemes. In contrast PNN along with feature reduction gives less computation time compared to the BPNN combination. Given the results, the proposed approach has a great potential for a variety of machine fault diagnostic applications as speed of computation by Artificial Neural Networks is superior.

REFERENCES