Classification of Machine Fault using Principle Component Analysis, General Regression Neural Network and Probabilistic Neural Network

M. F. Talib, M. F. Othman, N. H. N. Azli
Department of Electronic System Engineering, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia.
mdfauzi@utm.my

***Abstract*—As a major industry prime mover, induction motor plays an important role in manufacturing. In fact, production can cease its operation if there is some error or fault in the induction motor. In the industry, bearing, stator and rotor fault are the highest among other faults. Thus, this paper is to compare the accuracy of bearing, stator and rotor fault classification between General Regression Neural Network (GRNN) and Probabilistic Neural Network (PNN) with the previous work using Principle Component Analysis (PCA). The accuracy of fault classification for each method is improved by the selection of features extraction and number of classification. The features extraction used are mean, root mean square, skewness, kurtosis and crest factor. The sample data has been taken from Machinery Fault Simulator using accelerometer sensor, logged to text file using Labview software and analysed by using Matlab software. The accuracy of fault classification using GRNN method is higher than PNN because the sample data is classified through the regression of data as long as the sample data is redundant and lies on the regression distribution.**

***Index Terms*—GRNN; Machine Fault; PCA; PNN.**

I. INTRODUCTION

The Squirrel cage induction motors are widely used mostly in electrical machines for industrial, domestic and commercial applications. Different types of faults include stator winding faults, rotor bar breakage, misalignment, static or dynamic air-gap irregularities and bearing gearbox failures. The most common fault types of these rotating devices have always been related to machine shaft or rotor and bearing [1, 2]. Besides, the highest percentage fault in induction machine was bearing fault followed by stator, rotor and others which is around 40% to 45% as discovered by a survey done by the Institute of Electrical & Electronics Eng. and Electric Power Research Institute. In general, faults in electrical machines are dominated by failures in bearings and stator coils. For asynchronous motors with squirrel cage rotor, the failure statistics of bearings related fault is 41 percent while stator related faults is 37% followed by rotor faults which is 10% and other problems is 12% [3].

There are many types of fault in the bearing such as outer raceway fault, inner raceway fault and roller element fault [4], [5]. While in stator, the most common fault that happens is the breakdown of the winding insulation such as stator coils short-circuits and stator unbalance and eccentricities. Potential rotor faults in Brushless Direct Current (BLDC) motors machines are eccentricities and damaged rotor such as rotor broken bars and rings [6, 7]. Vibration signal analysis is a well-known and widely used diagnostic approach for bearing fault identification, and usually leads to good results in terms of effectiveness and detection capability [8, 10]. However, there are many types of monitoring that use advanced technologies in order to determine equipment condition and predict potential failure which are visual inspection, vibration measurement and analysis, temperature monitoring, acoustic emission analysis, noise analysis, oil analysis, wear debris analysis, motor current signature analysis, and non-destructive testing [11].

PCA is a classical statistical method for transforming attributes of a dataset into a new set of uncorrelated attributes called principal components (PCs). PCA can be used to reduce the dimensionality of a dataset, while still retaining as much of the variability of the dataset as possible. High dimensional data can pose problems for machine learning as predictive models based on such data run the risk of overfitting. Furthermore, many of the attributes may be redundant or highly correlated, which can also lead to the degradation of prediction accuracy [12, 13].

General Regression Neural Network (GRNN) is one of the most popular neural networks. GRNN is a feed-forward neural network for supervised data. It uses nonlinear regression functions for approximation. GRNN uses direct mapping to link the input layer to the hidden layer [14, 15]. Probabilistic neural networks can be used for classification problems.

PNN has the ability to train on sparse data sets. Moreover, it is able to classify data into specific output categories [16, 17]. There are a number of advantages of using PNN for classification. For example, the computational time of PNN is faster than BPNN, and it is more robust to noise. Furthermore, the training manner of PNN is simple and instantaneous [18].

II. METHODOLOGY

Generally, the method used to accomplish this experiment is described in

Figure **1**. It starts by acquiring a sample data from Machinery Fault Simulator (MFS). Then Matlab software is used to analyse the comparison of sample data using PCA, GRNN and PNN in order to see the accuracy of fault classification.

Figure 1: Flow chart

A. Data Acquisition

MFS is an innovative tool to learn and study the signatures of common machinery faults including bearing and induction motor defects.

Figure **2** shows the MFS that is used in the experiment.

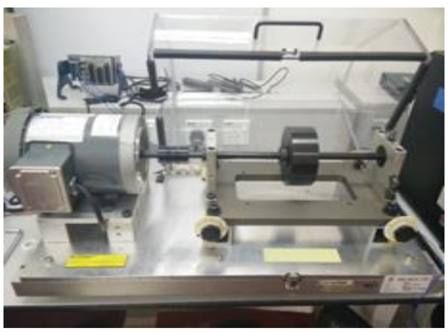

Figure 2: Machinery fault simulator

Figure **3** shows the flow chart to acquire the sample data. The experiment is setup at AC motor driver for the same frequency of 20 Hz for all sample data taken from an accelerometer sensor that is attached on MFS. NI cDAQ 9174 is used to get real time data sampling from the sensor. It is about 200,000 samples per test for 10 test of each bearing and induction motor. Labview software is used to log and save to text file before being analysed by Matlab. The types of sample data are as below:

1. Bearing
   a. Good Bearing
   b. Ball faulted bearing
   c. Outer race faulted bearing
   d. Inner race faulted bearing
   e. Combination faulted bearing
2. Induction motor
   a. Stator fault
   b. Rotor fault

B. Data Analysis

Seven types of sample data are acquired from MFS using Labview software and save to text file. Matlab software is used to analyse the sample data from text file. The parameter of sample data consists of acceleration in time-domain.

Figure **4** shows the flow chart of sample data analysis.

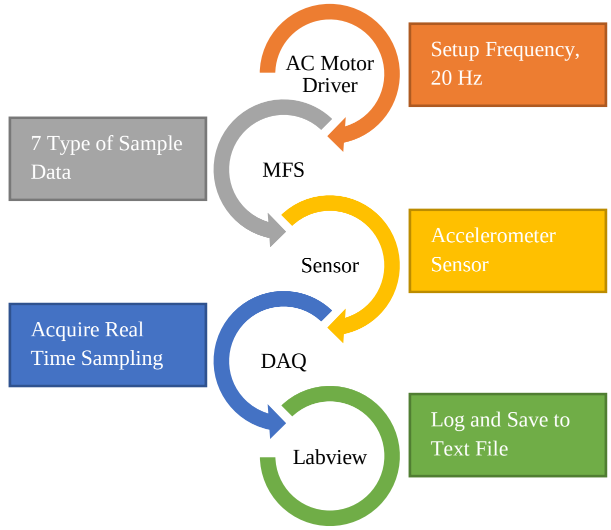

Figure 3: Flow chart to acquire the sample data

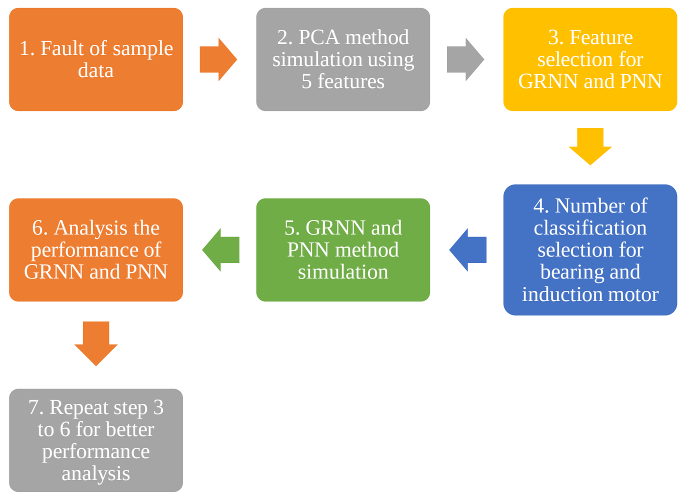

Figure 4: Flow chart of sample data analysis

The sample data are simulated using PCA method from previous work for the classification of each type of machine fault within Matlab software. The features of PCA method are descriptive statistics, which are kurtosis, root mean square (RMS), mean, crest factor, and skewness. The accuracy of fault classification is observed, analysed and improved using the method of artificial intelligence technique, which are GRNN and PNN. Besides that, the accuracy of fault classification is enhanced by the feature selection and number of classifications. Figure 5 shows the improved method for accuracy of fault classification.

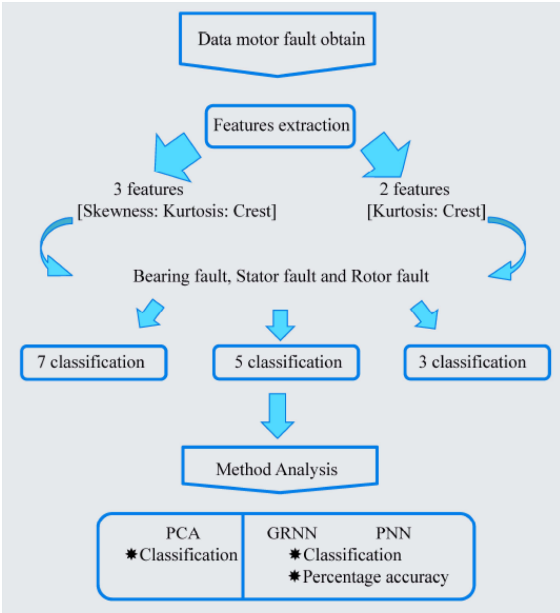

Figure 5: The improvement method for accuracy of fault classification

From five feature, only three and two features are used between Skewness, Kurtosis, and Crest because these features contribute different pattern data compared to mean and root mean square. This feature also contributes to more accuracy in the performance of the classifier in classifying the sample data to their classes. The number of classification is divided into seven, five and three classes between good bearing, ball faulted bearing, outer race faulted bearing, inner race faulted bearing, combination faulted bearing, stator fault and rotor fault. The result is analysed for the accuracy in classification with calculation percentage for GRNN and PNN.

III. RESULT AND DISCUSSION

The simulation result is shown for seven and three classifications for all types of motor fault which are bearing, stator and rotor faults. The simulation to enhance accuracy classification use three features which are skewness, kurtosis and crest. Then, data analysis is obtained from simulation result using PNN and GRNN because these methods can classify fault accurately, even though in large amount of data faults. This due to GRNN which uses direct mapping to link the input layer to the hidden layer and also uses nonlinear regression functions for approximation. On the other hand, PNN is an artificial neural network for nonlinear computing which adopts the Bayes optimal decision boundaries. For PCA method, the motor fault can be classified according to the classes trough the reference features extracted from the data motor fault obtained. PCA can be used to reduce the dimensionality of a dataset, while still retaining as much of the variability of the dataset as possible. Thus, when using the PCA method it is easy to understand the concept classification of motor fault.

A. Simulation Result Using PCA

Figure **6** shows the three classifications of good bearing, outer fault bearing and rotor faults. Here, the fault class is clearly seen from the pattern of data which is different for each type of data motor fault.

Figure **7** shows the five classifications for good bearing and bearing fault which are inner, outer, ball and combination fault bearing. Good bearing and combination fault bearing are clearly seen, however there are redundant sample data fault for inner, outer and ball fault bearing caused by the pattern of data being quite similar.

Figure **8** shows the result of seven classifications using PCA and the type of fault is differentiated by the type of colour such as red for healthy bearing, magenta for combination faulted bearing, blue for inner faulted bearing, green for outer faulted bearing, black for ball faulted bearing, clay for stator faulted and yellow for rotor faulted. From the result, there is overlap or redundant of some data motor fault such as inner, outer, ball faulted bearing, stator and rotor fault. The fault is classified according to its place. Thus, PCA can be used to improve the performance of machine learning methods in the classification of such high dimensional data. To overcome the fault or enhance the accuracy of these types of motor fault classification is by using artificial intelligence technique such as GRNN and PNN.

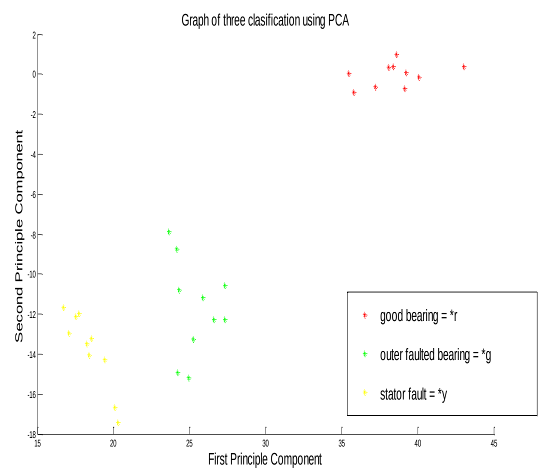

Figure 6: Three classification using PCA

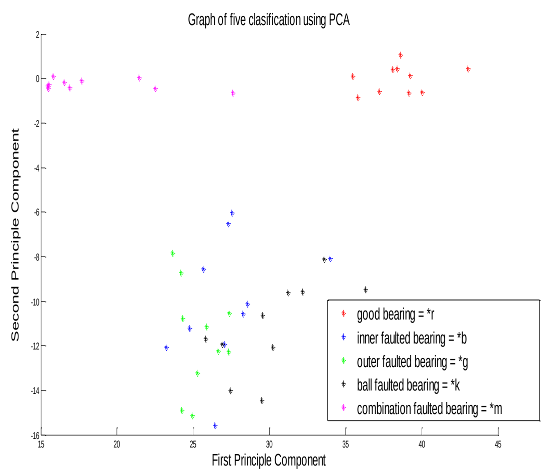

Figure 7: Five classification using PCA

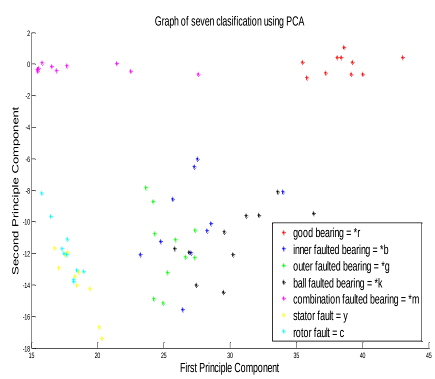

Figure 8: Seven classification using PCA

B. Simulation Result Using GRNN

Table 1 shows the result of classification machine fault for three classifications using GRNN which are good bearing, outer fault bearing and stator fault with two and three features selected and the percentage accuracy to determine the accuracy of performance classification of motor fault. While result percentage accuracy classification for this three machine fault shows a very good performance classification which is 100 percent accurate. Due to clear classification, it does not need to add new features into the simulation. Even when new features are added, the result is still the same.

Table 1
Three Classifications GRNN

Class	GRNN (%) 2 Features	GRNN (%) 3 Features
Good	100	100
Outer	100	100
Stator	100	100

Table 2 shows the result of classification motor fault for five classifications using GRNN which are good bearing, inner, outer, ball and combination fault bearing with two and three features selected and the percentage accuracy to determine the accuracy of performance classification of motor fault. The result of percentage accuracy classification for these five fault bearings is 100 percent accurate except for inner, ball and combination fault bearing which have low percentage accuracy which are 50, 60 and 90 percent respectively. Therefore, to improve the accuracy result, another feature is added into the simulation which is skewness where the percentage accuracy is improved from 10 to 30 percent for each type of motor fault classification. Besides, this result was obtained from 30 test sample data used as input compared to 70 test sample data that will disrupt the output classification accuracy result if this data fault is redundant or overlap each type of fault.

Table 2
Five Classifications GRNN

Class	GRNN (%) 2 Features	GRNN (%) 3 Features
Good	100	100
Inner	50	80
Outer	100	100
Ball	60	80
Combination	90	100

Table 3 shows the result of classification motor fault using GRNN which are good bearing, fault bearing and motor fault with two and three features selected which are skewness, kurtosis and crest factor and the percentage accuracy to determine the accuracy of performance classification of motor fault. While the result in percentage accuracy classification for motor fault approached to 100 percent accurate except for inner raceway fault bearing, roiling component fault bearing and rotor fault which have low percentage accuracy of 60 percent and below.

Table 3
Seven Classifications GRNN

Class	GRNN (%) 2 Features	GRNN (%) 3 Features
Good	100	100
Inner	50	80
Outer	100	100
Ball	60	80
Combination	90	100
Stator	100	90
Rotor	0	80

In case of GRNN, output is estimated using weighted average of the outputs of training dataset, where the weight is calculated using the Euclidean distance between the training data and test data. If the weight or distance is large then the weight will be very less and if the distance is small, it will put more weight to the output. Therefore, to improve the accuracy result, another feature is added into the simulation which is skewness that results the percentage accuracy to improve from 10 to 80 percent for each type of motor fault classification.

C. Simulation Result Using PNN

Table 4 shows the result of classification motor fault for three classifications only which are good bearing, outer fault bearing and stator fault with two and three features selected and the percentage accuracy to determine the accuracy of performance classification of motor fault. Based on the result, the percentage accuracy of these three classes of fault is higher compared to other result which approaches 100% accuracy. This is due to the fact that PNN can clearly and easily classify the fault according to its class. However, the accuracy is reduced after adding a new feature which is skewness.

Table 4
Three Classifications PNN

Class	PNN (%) 2 Features	PNN (%) 3 Features
Good	100	100
Outer	80	80
Stator	100	70

Table 5 shows the result of classification machine fault for five classifications which are good bearing, inner, outer and ball and combination fault bearing with two and three features selected and the percentage accuracy to determine the accuracy of performance classification of motor fault. The result of percentage accuracy classification for this five machine fault is low performance classification which is not more than 60% accuracy. This is due to unclear segment cases classification because this three data motor fault is redundant to each other, causing the classifier to exhibit good output classification.

Table 5
Five Classifications PNN

Class	PNN (%) 2 Features	PNN (%) 3 Features
Good	100	100
Inner	30	20
Outer	40	20
Ball	20	0
Combination	70	70

Table 6 shows the result of classification motor fault for seven classifications using PNN which are good bearing, fault bearing, and motor fault with two and three features selected and the percentage accuracy to determine the accuracy of performance classification of motor fault. The percentage in accuracy classification for this seven motor fault shows to be a good performance classification because most of the motor fault does not reach 100 percent accuracy. This is due to the fact that redundant samples can potentially lead to a large network structure that causes the classifier to be oversensitive to the training data and is likely to exhibit poor generalization capacities to the unseen data.

Table 6
Seven Classifications PNN

Class	PNN (%) 2 Features	PNN (%) 3 Features
Good	100	100
Inner	40	20
Outer	40	20
Ball	20	30
Combination	60	70
Stator	100	80
Rotor	10	10

Thus, there is an outstanding issue associated with PNN concerning network structure determination, which is determining the network size, locations of pattern layer neurons as well as the value of the smoothing parameter.

D. Comparison between GRNN and PNN

From Tables 1 and 4, the percentage accuracy with two and three features using GRNN is higher than the percentage accuracy using PNN. This is due to the sample data fault being in clear boundary position, thus producing very good performance classification for both methods because the percentage accuracy almost approaches 100 percent. Other than that, the size of testing sample data is small that causes pattern recognition easily done by the classifier.

From Tables 2 and 5, the percentage accuracy with two and three features using GRNN is higher than percentage accuracy using PNN. This due to the sample data fault is in clear boundary position thus producing very good performance classification even though these three data fault bearings overlap with others for both methods because the percentage accuracy almost reaches 100 percent. This is due to the feature selected had different distribution data that causes GRNN to still classify the bearing fault according to its class. Other than that, the size of testing data sample is small that causes pattern recognition to be easily done by the classifier.

From Tables 3 and 6, the percentage accuracy with two and three features using GRNN is higher than percentage accuracy using PNN. This is due to the sample data fault is in line regression, thus the classifier will exhibit excellent generalization output of sample data fault result.

IV. CONCLUSION

This section describes the summary of results obtained by using three methods, which are PCA, GRNN and PNN. PCA is a classical statistical method for transforming attributes of a dataset into a new set of uncorrelated attributes. PCA can be used to reduce the dimensionality of a dataset, while still retaining as much of the variability of the dataset as possible. High dimensional data can pose problems for machine learning as predictive models based on such data run the risk of overfitting. GRNN is a feed-forward neural network for supervised data. It uses nonlinear regression functions for approximation. GRNN uses direct mapping to link the input layer to the hidden layer. While, PNN is an artificial neural network for nonlinear computing which approaches the Bayes optimal decision boundaries.

This is done by estimating the probability density function of the training dataset using the Parzen nonparametric estimator. Bayesian strategies are decision strategies that minimize the expected risk of a classification. From the result obtained using PNN and GRNN, it shows that the GRNN method is higher than PNN because the sample data is classified through the regression of data as long as the sample data is redundant and lies on the regression distribution. The classifier can classify the data of motor fault according to its class. The accuracy for classification using GRNN can be enhanced by adding another features while for the analysis method using PNN, there is no improvement although new features are added. Next, analysis method using GRNN can still classify the fault according to the type of fault such as bearing fault, stator fault and rotor fault even though there is seven class but analysis method using PNN cannot classify the fault when the input was more than three. Therefore, the percentage accuracy when using GRNN is higher than the percentage accuracy when using PNN.

ACKNOWLEDGMENT

The authors would like to thank the Malaysia-Japan International Institute of Technology (MJIIT) and Universiti Teknologi Malaysia (UTM) under Grant No 4B212 for their support.

REFERENCES

[1] Castelli Marcelo, Juan Pablo Fossatti and Jose Ignacio Terra, "*Fault Diagnosis of Induction Motors Based on FFT*" Fourier Transform - Signal Processing, InTech. Uruguay, Universidad de Montevideo, 2012

[2] Khadim Moin Siddiqui, Kuldeep Sahay, V.K.Giri, "Health Monitoring and Fault Diagnosis in Induction Motor- A Review," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 3, no. 1, pp. 6549–6565, 2014.

[3] Fernando Briz, Michael W. Degner, Pablo García, David Díaz-Reigosa and Alberto Díez, "*Rotor Fault Detection in Line-fed Induction Machines Using Complex Wavelet Analysis of Startup Transients*." University of Oviedo, Ford Motor Company Spain, USA, 2010.

[4] Dr. Shashi Raj Kapoor, Nishant Khandelwal and Parikshit Pareek, "Bearing Fault Analysis by Signal Energy Calculation Based Signal Processing Technique in Squirrel Cage Induction Motor," *Signal Propagation and Computer Technology (ICSPCT), 2014 International Conference on*. Ajmer IEEE, 2014, pp. 33–38

[5] Martin Bl¨odt, Pierre Granjon, Bertrand Raison, Gilles Rostaing, "*Models for Bearing Damage Detection in Induction Motors Using Stator Current Monitoring*" IEEE Transactions on Industrial Electronics, Institute of Electrical and Electronics Engineers, vol. 55, no. 4, pp. 1813–1822, 2008.

[6] M. H. Song, E. Kang, C.H. Jeong, M.Y. Chow & B. Ayhan, "Mean Absolute Difference Approach for Induction Motor Broken Rotor Bar Fault Detection" *Symposium on Diagnostics for Electric Machines, Power Electronics and Drives Atlanta,* CA, USA. 2003, pp. 115–118.

[7] Hua-Qing Wang, Wei Hou, Gang Tang, Hong-Fang Yuan, Qing-Liang Zhao and Xi Cao. "Fault Detection Enhancement in Rolling Element Bearings via Peak-Based Multiscale Decomposition and Envelope Demodulation," *Mathematical Problem in Engineering*, pp. 1–11, 2014.

[8] Akshat Singhal and Meera A.Khandekar, "Bearing Fault detection in induction motor using motor current signature analysis" *International Journal of Advanced Research in Electrical, Electronic and Instrumentation Engineering*, vol. 2, no. 7, pp. 3258–3264, 2013

[9] Latuny, Jonny, "*A sensitivity comparison of Neuro-fuzzy feature extraction methods from bearing failure signals*" Doctor of Philosophy of Curtin University, Department of Mechanical Engineering, 2003.

[10] Pratesh Jayaswa, A. K. Wadhwani, and K. B. Mulchandani, "Machine Fault Signature Analysis" *International Journal of Rotating Machinery*, Hindawi Publishing Corporation, vol. 2008, pp. 2–10, 2008

[11] Bostjan Dolenc, Pavle Boskoski and Dani Juricic, "Distributed bearing fault diagnosis based on vibration analysis" *Mechanical Systems and Signal Processing*, vol. 66–67, pp. 521–532, 2015.

[12] Ryder, A., "Classification of narcotics in solid mixtures using Principal Component Analysis and Raman spectroscopy and chemometric methods" *Journal of Forensic Sciences*, vol. 47, no. 2, pp. 275–284, 2002.

[13] O.R. Seryasat, H. Ghayoumi Zadeh, M. Ghane, Z. Abooalizadeh, A. Taherkhani, F. Maleki, "Fault Diagnosis of Ball-bearings Using PCA and Support-Vector Machine" *Life Science Journal 2013*, vol. 10, no. 1s, pp. 393–397, 2013.

[14] Ahmed Amara Konate, Heping Pan, Nasir Khan, Jie Huai Yang, "Generalized regression and feed-forward back propagation neural networks in modelling porosity from geophysical well logs" *Journal of Petroleum Exploration and Production Technology,* vol. 5, pp. 157–166, 2015.

[15] Buse Melis Ozyildirim, Mutlu Avci, "*Generalized classifier neural network*" Department of Computer Engineering University of Adana Science and Technology, Adana, Turkey, Neural Networks, vol. 39, pp. 18–26, 2013.

[16] Abdelhadi Lotfi and Abdelkader Benyettou, "A reduced probabilistic neural network for the classification of large databases" *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 22, no. 4, pp. 979–989, 2014.

[17] Abdelhadi Lotfi and Abdelkader Benyettou, "Using Probabilistic Neural Networks for Handwritten Digit Recognition" *Journal of Artificial Intelligence*, vol. 4, pp. 288–294, 2011.

[18] T. Santhanam, S. Radhika, "Probabilistic Neural Network - A Better Solution for Noise Classification" Journal of Theoretical and Applied Information Technology, vol. 27, no. 1, pp. 39–42, 2011.