International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-02, MAY 2021

Fault Classification of Industrial Motor using Random Forest Algorithm

Reshma K L, M.Tech Scholar, RSET, Kochi, India, reshmamary888@gmail.com

Jobin K Antony, Professor, RSET, Kochi, India, jobinka@rajagiritech.edu.in

Abstract- **Large rotating machines are widely used at factories, ships and power plants and each rotating machine vibrates. If the vibrations are excessive, they can pose problems to the usage of the machine and reduce its lifetime. The use of vibration analysis can determine problems caused due to improper installation, machining errors, insufficient lubrication, improper shaft or sheave alignment, loose bolting, bent shafts, and much more. Machine learning can be introduced to predict the type of fault of the machine according to the frequency of the vibrations. The vibrations were measured using an accelerometer and stored in a database. This data can be used for analysis in both time domain and frequency domain.**

Keywords — Accelerometer, FFT, Machine Learning, Random Forest Classifier, Raspberry-Pi, Vibration Analysis

I. INTRODUCTION

A rotating machine or a motor is a very commonly used industrial equipment. Every rotating object will have some vibrations associated with it. Even though small vibrations of a machine are acceptable, if it exceeds certain limits, it is an indication that there is a probable deviation from its normal operations. These slight changes in the vibration pattern may not be noticed by a maintenance team in an industry, which will even worsen the motor condition due to negligence [6]. The motor vibrations best describe the health of a rotating machine; hence vibration characteristics was taken for the analysis. The incoming vibration can be analyzed using various methods. The raw vibration data can be taken for the time domain analysis, which will measure the peak amplitude, rms value etc. These results can be used to monitor the motor conditions, but it gives only a less amount of information about the motor health. Another method is to perform the analysis in the frequency domain. This can be achieved by performing the Fourier transform on the time domain vibration signal. The frequency domain analysis provides more useful results because the frequency peaks obtained can be used to differentiate the motor states.

By introducing the machine learning field into this analysis, it becomes easier to perform an analysis based on the vibration data sensed from the motor. A machine learning algorithm can be used to classify the fault of the rotating machinery after training the model.

In this paper, the motor vibrations were sensed using a MEMS accelerometer which was attached to the machine. The motor vibrations were taken and sent to a Raspberry-pi board [7], which sent the accelerometer and gyroscope readings to a cloud database. This allows real time monitoring of the data from anywhere. These readings were then analyzed in the time and frequency domain. Using a machine learning algorithm, the motor health was classified as normal and faulty based on the frequency peaks determined from the Fast Fourier Transform (FFT).

II. HARDWARE SETUP

This project uses an embedded platform and an MPU-6060 MEMS accelerometer which has a 3-axis accelerometer and gyroscope.

For hardware setup as shown in Fig.1, the sensor was connected to the raspberry-pi board using the I^2C communication protocol. For the connection of the sensor to the raspberry-pi board, the SCL, SDA pins of the MPU-6050 are used. The pi-board can then be connected to a laptop using the ethernet cable or through wi-fi network.

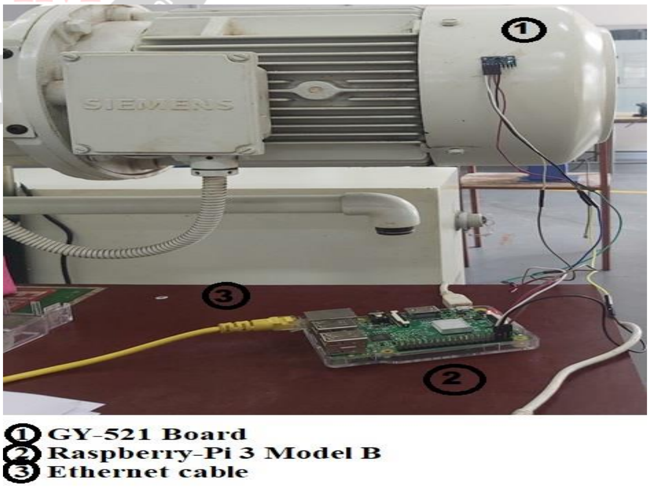

Fig.1: Hardware Setup

III. SOFTWARE SETUP

The following software setups should be done: -

1. Setup the raspberry-pi board by installing the Raspbian OS.

 DOI : 10.35291/2454-9150.2021.0231

2. Determine the device address to interface the sensor module to the pi-board.
3. Determine the IP address of the raspberry-pi using the "Advanced IP scanner".
4. Use the "putty" software and the "VNC viewer" to view the raspberry-pi in a laptop.
5. The sensor values read can be written into a google sheet that is already created.

IV. DATA ANALYSIS

A. Time Domain Analysis

In this method, the vibration data was plotted against time. This is the simplest form of analysis.

Some of the features that can be determined from a plot in time domain includes amplitude- it provides the maximum and minimum value of acceleration or vibration of a motor. Another quantity that can be measured is the rms value. This measurement helps to determine the intensity of the vibrations obtained. Peak to peak value gives the range of the vibration signal.

Root Mean Square.

This measurement helps to determine the intensity or energy of the vibrations obtained. The RMS value of the vibrations obtained from a faulty bearing is more than that of a healthy bearing.

$$Y_{RMS}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}$$

Kurtosis.

Kurtosis shows the occurrence of sharp peaks in the signal. The kurtosis value of a normal bearing is less than 4. A defective motor bearing has a high kurtosis value between 4 and 45.

$$\text{Kurtosis}=\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(x(i)-mean)^4}{(standard\ deviation)^4}}$$

Crest Factor.

It is the ratio of RMS and maximum value of a vibration signal.

$$\text{Crest Factor}=\frac{x_{peak}}{x_{RMS}}$$

Even though the time domain analysis is simple and easy, it does not give much useful information to determine the health of a machinery. Most of the vibration signals in the time domain are not clearly differentiable. Hence, we go for analysis in the frequency domain.

B. Frequency Domain Analysis

Frequency analysis [9] is a more accurate method for monitoring the condition of machines. It provides information based on frequency characteristics that are not easily observed in the time domain.

The frequency peaks obtained corresponds to the state of the machinery. Therefore, an abnormal frequency peak found out while plotting a frequency spectrum corresponds to an abnormal state of a machinery. i.e., it means that the machine is at fault. This makes the fault detection and classification easier.

Some of the limitations in using frequency domain analysis are- it cannot be used to differentiate the spectrum if the input signals are varying over time. i.e., input signals should be stationary.

Another limitation of a spectral analysis method is that, it cannot detect small spikes in the signal. It was seen that, although we expected the spectrum of a signal with and without small transients in the input signal to be different, the plots were the same.

To overcome these limitations, we can consider the frequency change of the signal with respect to time. For that, short time Fourier transform can be performed, which is a time-frequency analysis method.

V. MACHINE LEARNING

A. Machine Learning Types

Today, artificial intelligence and machine learning have shown their influence in every field of applications. A machine learning algorithm can be used to analyze the data obtained from the sensor and determine the type of fault of the machine inspected [4]. There are mainly three classes of machine learning which includes supervised learning, unsupervised learning and reinforcement learning.

Supervised learning requires an input and a corresponding label. In this method, during the training phase, the inputs are mapped to the corresponding output. Therefore, both the input and the corresponding outputs should be known at the time of training. Some of the supervised learning algorithm includes Support Vector Matrix (SVM), Random Forest (RF), Linear Regression etc.

The next class of machine learning algorithm comes under the category of unsupervised learning. They do not require an output label during the training phase. In this type, the input data is only taken during the training phase and the inputs will itself form a group or cluster according to the common characteristics among them. Some of the unsupervised algorithms include K-means clustering algorithm, etc.

The next type of machine learning algorithm is called the reinforcement learning. It also does not require an output label during the training phase. Only the input data known as an 'agent' is used. In this learning method, the input agent is trained in an environment and they are given reward points (positive for correct and negative for wrong). The agent tries to gain maximum reward points and by a trial-and-error method, it will eventually learn and will have

 DOI : 10.35291/2454-9150.2021.0231

ISSN : 2454-9150 Vol-07, Issue-02, MAY 2021

maximum rewards which means that it will give accurate results. Q-learning algorithm is an algorithm based on reinforcement learning.

In this paper, the random forest algorithm, which is a supervised learning method was used.

B. Random Forest Algorithm

It is based on decision trees.[2]

In this method, the original training dataset is randomly divided into several subsets that can be used to induce multiple classifiers. Then, a combination process is used to produce a single classification for a given example.

The basic idea of this method is that for the k^{th} tree, a random vector ν^k is produced, independent of the past random vectors ν_1, ν_2, ν_3, …, $\nu_k - 1$ but with the same distribution. With these generated random vectors, the generated k trees are used to vote for the most popular class label. These processes are called random forests (RFs) [8]. In other words, an RF is an ensemble learning method that constructs multiple Decision Trees during the training process; the final output is the mode of the classes output by each individual Decision Tree.

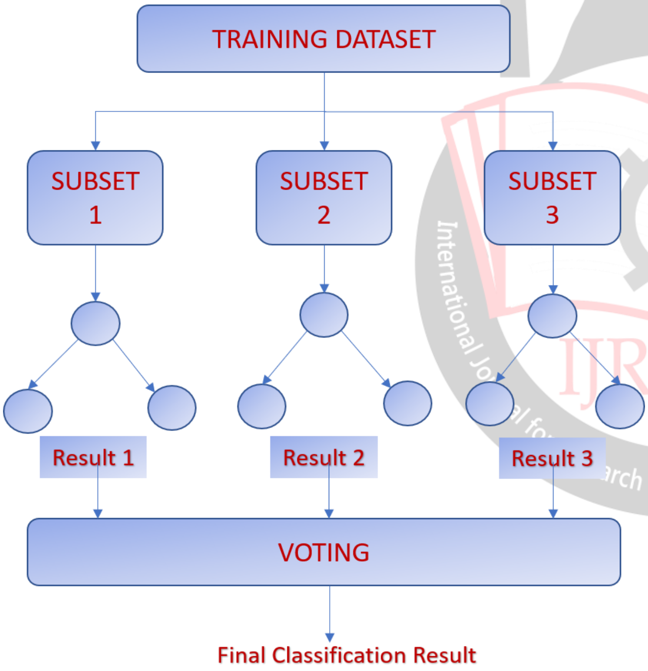

Fig. 2. Random Forest Classifier

As shown in Fig.2, in a random forest classifier, the training dataset is first divided into different subsets. Then each subset is trained individually similar to a decision tree classifier. After that, the result obtained from each individual tree subset is taken and the final classification result is calculated.

C. Data Collection

The first step in any machine algorithm is data collection. More the data, more accurate will be the results. In this paper, the dataset was taken from an online database that is freely available. i.e., the case western university bearing data set [5].

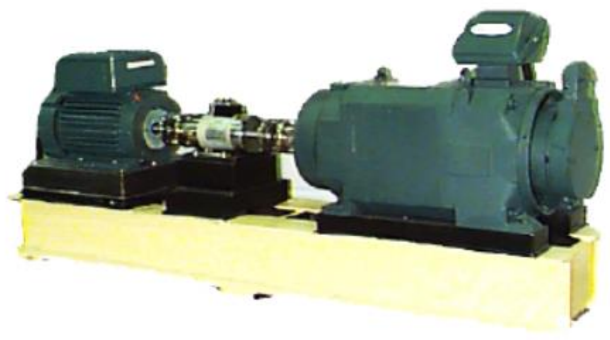

Fig.3. Test stand for Case Western Reserve University bearing dataset

The test stand used to acquire the Case Western Reserve University (CWRU) bearing dataset is illustrated in Fig.3, in which a 2-hp induction motor is shown on the left, a torque transducer/encoder is in the middle, while a dynamometer is coupled on the right. Single point faults are introduced to the bearings under test using electro-discharge machining with fault diameters of 7 mils, 14 mils, 21 mils at the inner raceway, the rolling element and the outer raceway. Vibration data are collected for motor loads from 0 to 2 hp and motor speeds from 1730 to 1,797 rpm using two accelerometers installed at both the drive end and fan end of the motor housing, at sampling frequency of 12 kHz. The generated dataset is recorded and made publicly available on the CWRU bearing data center website.

D. Feature Extraction

In this step, a new dataset is taken from the given dataset. For this project, we performed the Fourier transform on the database available. From the Fourier transform obtained, the peak frequency was found out. This peak frequency was selected as the 'feature' for the algorithm [4]. After determining the peak, it was mapped to the corresponding state of the machinery. This was done as we are using a supervised learning algorithm, which require the input and output to be mapped during the training phase. Now we have prepared the training dataset. From the total available dataset, 80 percent was taken for training and the remaining 20 percent was taken for testing.

VI. PROPOSED METHOD

This paper aims to provide a real time fault classification of an industrial machinery. The project is mainly divided into two parts: - data acquisition and data analysis.

The data acquisition part involves connecting the sensor to a motor and reading the vibration values. The acquired readings are stored into an online database with the help of the raspberry-pi board.

 DOI : 10.35291/2454-9150.2021.0231

The data analysis part involves analyzing the data acquired in time domain and frequency domain and using this information to determine the health of the machine.

The following steps were carried out: -

- Complete the hardware and software setup as discussed in the previous sections.
- Read the vibration data available from the bearing dataset.
- Determine the FFT of the signal.
- Determine the peak frequency.
- Create a Data-frame in python for different test cases. i.e. for normal and faulty readings. The health of the motor corresponds to the frequency peaks obtained in the previous step.
- Perform training and testing using a Random Forest Classifier.

VII. RESULT

The vibration signals were analyzed in both time domain and frequency domain. Figures show the plot of vibration signals in the time domain and its corresponding FFT plot.

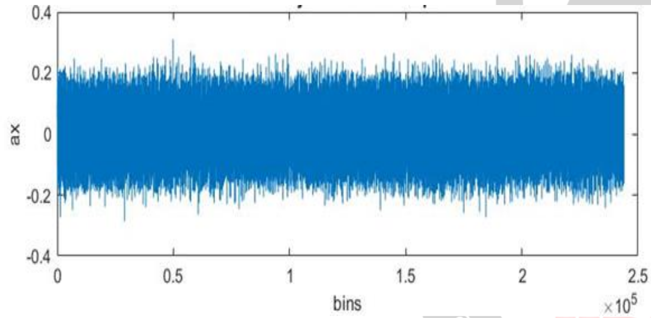

Fig. 1. Time domain plot of a healthy bearing

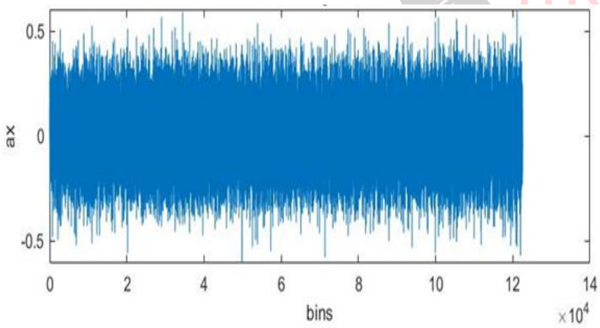

Fig. 2. Time Domain plot of a faulty bearing

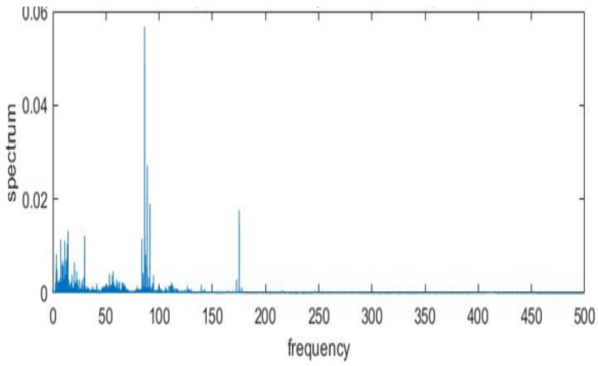

Fig. 3. Frequency spectrum of healthy bearing

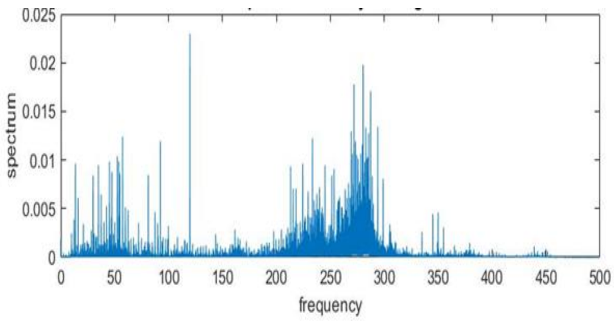

Fig. 4. Frequency spectrum of faulty bearing

Fig. 3 and Fig.4 shows the vibration signal of a healthy and faulty motor in the time domain.

As we can see, the amplitude of the signals is different for the two, time domain plots, but it is difficult to differentiate the motor state according to these vibration patterns. Hence time domain analysis is not good for determining the fault of the machine.

Fig. 5 and Fig.6 shows the state of the same motors in the frequency domain by plotting the FFT.

These plots are significantly different in their frequency domain; hence it is easier to differentiate a machine that is healthy and faulty. We can see that, the FFT of the motor with fault shows many abnormal frequency peaks, which shows that the machine is at fault, thereby making the vibration analysis easier.

Sl. No.	Type of fault	Speed (rpm)	Location	% Accuracy		
				RF	DT	k-NN
1	Outer Raceway	1,797	Drive End	100	100	100
2	Outer Raceway	1,772	Drive End	90	100	70
3	Outer Raceway	1,750	Drive End	60	100	90
4	Outer Raceway	1730	Drive End	100	100	90
5	Inner Raceway	1,797	Drive End	100	100	100
6	Inner Raceway	1,772	Drive End	100	100	100
7	Inner Raceway	1,750	Drive End	100	70	90
8	Inner Raceway	1,730	Drive End	100	80	60
9	Ball Defect	1,797	Drive End	100	100	100
10	Ball Defect	1,772	Drive End	60	0	30
11	Ball Defect	1,750	Drive End	100	100	100
12	Ball Defect	1,730	Drive End	100	100	100
13	Healthy	1,797		100	90	0
14	Healthy	1,772		30	40	100
15	Healthy	1,750		90	0	100
16	Healthy	1,730		90	90	0
17	Ball Defect	1,797	Fan End	100	0	90
18	Ball Defect	1,772	Fan End	90	40	30
19	Ball Defect	1,750	Fan End	100	50	90
20	Ball Defect	1,730	Fan End	100	100	60
21	Outer Raceway	1,797	Fan End	100	100	80

 DOI : 10.35291/2454-9150.2021.0231

22	Outer Raceway	1,772	Fan End	100	100	80
23	Outer Raceway	1,750	Fan End	100	100	60
24	Outer Raceway	1,730	Fan End	100	50	50
25	Inner Raceway	1,797	Fan End	100	100	100
26	Inner Raceway	1,772	Fan End	100	100	60
27	Inner Raceway	1,750	Fan End	100	100	70
28	Inner Raceway	1,730	Fan End	100	100	100
	Total accuracy (%)			**93.2**	**82.5**	**75**

Table I. Accuracy of various algorithms

The Table.I. shows a comparison of the accuracy obtained by the fault classifier by using various machine learning algorithms- Random Forest (RF), Decision Tree (DT) and K-Nearest Neighbors (k-NN). It was observed that the Random Forest algorithm performed the fault classification with the highest accuracy.

Fig.7 shows the confusion matrix for a randomforest classifier. Confusion matrix is an evaluation matrix used on machine learning which will show the true positive and true negative values. Here the confusion matrix is created by taking twelve classes, in which one class is of the healthy motor and the others correspond to different type of motor faults. The diagonol of the confusion matrix shows the true positive value, which means that the true values and the predicted values are equal.

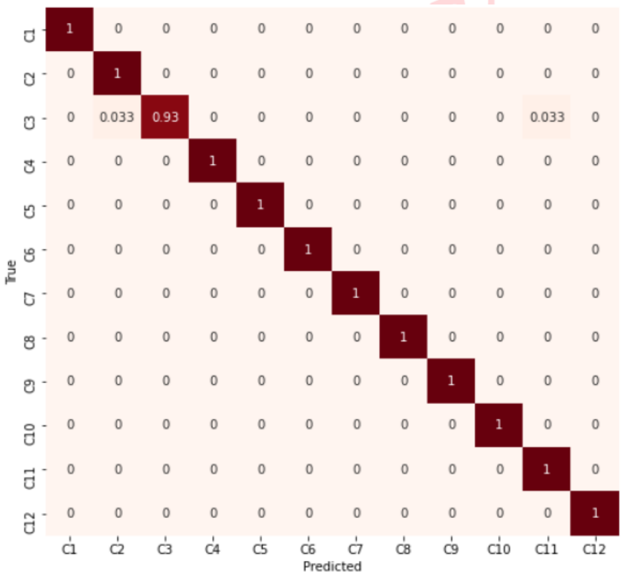

Fig. 7. Confusion Matrix of random forest classifier

VIII. FUTURE WORKS

Predictive maintenance is a technique that uses condition monitoring tools and techniques to track the performance of an equipment during normal operation to detect possible defects and fix them before they result in failure. The ultimate goal of this approach is to perform maintenance at a scheduled point in time when the maintenance activity is most cost-effective and before the equipment loses performance within a threshold. This results in a reduction in unplanned downtime costs.

When predictive maintenance is working effectively, maintenance is only performed on machines when it is required. That is, just before failure is likely to occur. The frequency peaks of the vibration signals can be analyzed so that it can detect the peaks that are near to the peaks of a faulty machine. This helps in performing the necessary maintenance to the machine before the actual fault occurs.

A machine learning problem requires large computational resources for implementing the algorithm and also requires large storage capacity for storing the big-data. Therefore, hardware optimizations have now evolved. i.e., the machine learning algorithm are being implemented on a hardware, after performing all the software optimizations.

Some of the techniques that can be used for hardware acceleration are by using Graphical Processing Unit (GPU), Application Specific Integrated Circuits (ASIC) or by using a Field Programmable Gate Array (FPGA). Among these, an FPGA is preferred due to its reconfigurability, large storage space etc. This project can be extended by implementing the random forest classifier on an FPGA board for hardware acceleration.

IX. CONCLUSION

The main objective of this project known as fault classification, was to perform classification of fault in an industrial machinery to determine the health of the machine. For that, vibration analysis was performed on the machine. The vibration signals from the machine were sensed using an accelerometer which was mounted on the motor and using an embedded board, the readings were stored into a database. From the csv files which has the vibration readings, the signals were plotted in both time and frequency domain. The frequency domain signals were better to differentiate the vibration data for healthy and fault machine. Then a machine learning algorithm was used for fault classification and prediction.

The dataset available from the Case Western Reserve University bearing dataset were used for creating the training dataset for the fault classification. The frequency domain analysis was done to determine the feature to be used for the training purpose. A random Forest classifier was used as the machine learning algorithm. It classified the incoming dataset into healthy machine or a machine with fault according to the frequency peaks obtained by performing the FFT on the incoming vibrations.

Three different faults were identified, i.e., inner raceway fault, outer raceway fault and rolling element fault at different motor speeds and motor loads. The accuracy of this model was found to be around 93%.

REFERENCES

[1] S. Zhang, S. Zhang, B. Wang and T. G. Habetler, Deep Learning Algorithms for Bearing Fault Diagnostics A

 DOI : 10.35291/2454-9150.2021.0231

Comprehensive Review-*IEEE Access*, vol. 8, pp. 29857-29881,2020.

[2] Asoke K. Nandi, Hosameldin Ahmed, "Decision Trees and Random Forests," in Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines," *IEEE*, 2019, pp.199-224

[3] A. Alghassi, Z. Yu and F. Farbiz, "Machine Performance Monitoring and Fault Classification using Vibration Frequency Analysis," 2020 Prognostics and Health Management Conference (PHM Besanon), Besancon, France, 2020, IEEE, pp. 8-14, doi: 10.1109/PHMBesancon49106.2020.00009.

[4] Brandao, Iago, "Vibration Analysis of Rotary Machines Using Machine Learning Techniques", 2019, European Journal of Engineering Research and Science.

[5] Case Western Reserve University, Seeded Fault Test Data, http://csegroups.case.edu/bearingdatacenter/home

[6] Ivar Koene, Raine Viitala, Petri Kuosmanen, "Internet of Things based monitoring of Large Rotor Vibrations with a Microelectromechanical systems accelerometer", - IEEE Access Vol 7, 2019

[7] Marek Iwaniec, Andriy Holovatyy, Vasyl Teslyuk," Development of Vibration Spectrum Analyzer using the Raspberry-Pi Microcomputer and 3-Axis Digital MEMS Accelerometer ADXL345", MEMSTECH 2017, 20-23 April, 2017, Polyana-Svalyava (Zakarpattya), UKRAINE

[8] Breiman, L, "Random forests- Machine Learning," 45 (1): 532. Springer.

[9] Shunming Li1, Yu Xin1, Xianglian Li2, Jinrui Wang1, Kun Xu1, A Review on the Signal Processing Methods of Rotating Machinery Fault Diagnosis, 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC 2019)

[10] Sal Langarica, Christian Rffelmacher, and Felipe Nez An Industrial Internet Application for Real-Time Fault Diagnosis in Industrial Motors, IEEE Transactions On Automation Science And Engineering, April 19, 2019 .

 DOI : 10.35291/2454-9150.2021.0231