

**Predictive Maintenance on Bearing for Industry 4.0**

**Introduction**

Predictive maintenance techniques are being developed to help analyse the condition of in-service equipment in order to estimate when maintenance should be conducted. This method saves money over routine or time-based preventative maintenance because actions are only performed when necessary. As a result, it is considered as condition-based maintenance carried out in accordance with an equipment’s deteriorating state estimates. Data-driven predictive maintenance is very beneficial and successful since it reduces downtime and helps to optimise spare parts inventories. Preventive maintenance is essential for extending the life of equipment. A simulation of the bearing machinery may be created, which subsequently allows us to create datasets from the simulation model. The dataset includes both functional and incorrect data from the bearing machines. We can improve the data from the physical machines by using these synthetic datasets.

**Literature Survey and Review**

In 2017[a], the work “Fault diagnosis of roller bearings using parameter assessment technique and multiclass support vector machine” was published in the AIP Conference Proceedings Journal. It stands for Parameter Evaluation Technique in SVM classification, which aims to obtain the highest level of accuracy in determining the type of defect that occurs in bearing machines as well as predicting the overall health of the bearings in the machines. Multiple research articles and publications including the condition monitoring of machinery such as compressors and power converters using MATLAB were published in Science Direct, IEEE, and Intech Open Journals in the same year. In 2019, the AIP Conference Proceedings Journal published an article titled "Ball bearing defect diagnostics using wavelet transform and principal component analysis"[b]. The wavelet transform and PCA for obtained vibration signals for roller bearings were explained in this study. Multiple papers titled “Acoustic Fault Analysis of Three Commutator Motors”[c] and “Bearing Fault Diagnosis Based on VMD-SVD and Fuzzy Clustering”[d] were published in Science Direct Journal and International Journal of Pattern Recognition and Artificial Intelligence in 2019 involving Fault Detection using K-means Clustering. These study papers explained fault diagnosis using a mix of K-means and Fuzzy Clustering, which employs decomposition methods to find the most accurate fault characteristics of bearings. We tracked the development of methods and algorithms in the field of fault diagnosis in machinery that uses bearings, such as simple bearings, roller bearings, and ball bearings. In the realm of machine maintenance plans including bearings, the following flow diagram depicts how technology and algorithms for defect identification have advanced over the last four years (2017-2020). The evolution of Fault Detection Strategies is depicted in Figure 4.

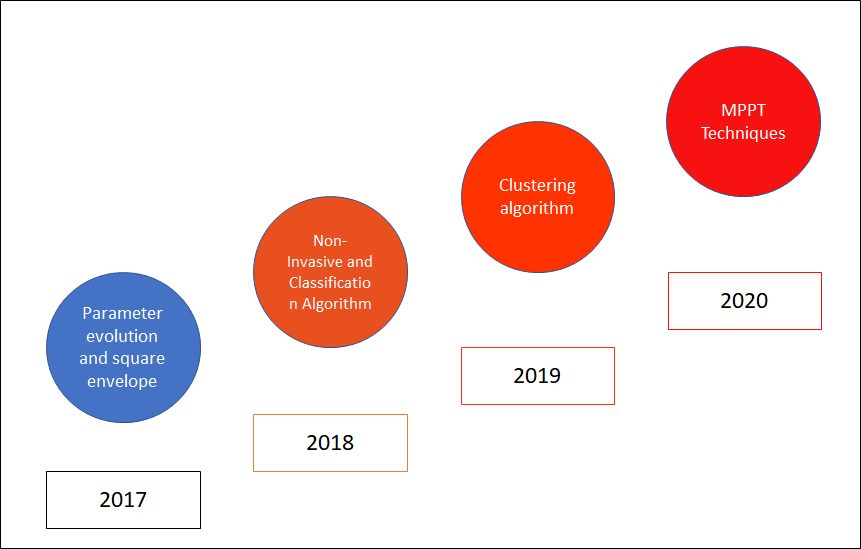


Fig 2: Evolution in Fault Detection Strategies After complete research and review,

After thorough investigation and analysis, it can be concluded that during the operation of mechanical systems, the machine is normally accompanied by strong nonlinear and non-stationary vibrations, and the failure of the rolling bearing would further cause unwanted vibration of other components, resulting in the measured vibration signals being mixed with or submerged in the vibration signals of other parts[e]. This necessitates the detection and diagnosis of these defects in order for all mechanical components of a machine to function properly.

**Technologies Used**

Python3, Google Colaboratory, Jupyter Notebook, Tensorflow 2.6.0, Keras 2.6.0, matplotlib 3.2.2, numpy, pandas, scikit-learn, scipy, Self Organizing Maps Package sompy 1.1.1.

**Dataset Description**

The Prognostic Health Maintenance (PHM) dataset was focused on the estimation of remaining useful lifetime of bearings. It was provided by FEMTO-ST Institute[7]. Pronostia is an experimental platform to test and validate bearings fault detection, diagnostic and prognostic approaches. It is composed of three main parts: a rotating part, a loading part and a measurement part.

Rotating part: We can operate the machine interface to set the speed, select the direction of rotation and set the monitoring parameters.

Loading part: It works on the dynamic load. The load is generated by the force actuator.

Measurement part: Operating conditions are measured by the instantaneous measures of the radial force on the bearing. Vibration sensors will have two accelerometers placed at 90 degrees to each other on the vertical and horizontal axis.

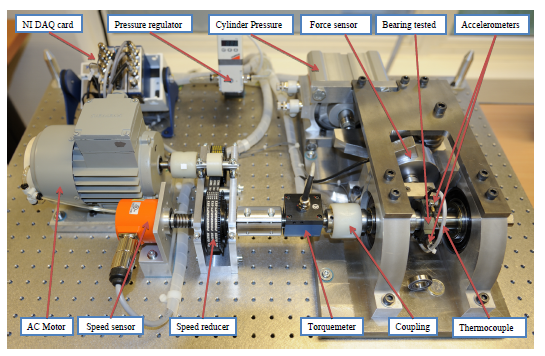


Fig 3: Overview of PRONOSTIA [9]

Used Dataset: The data consists of three different loads:

1. Conditions 1(First Operating conditions): 1800 rotations per minute(rpm) and 4000 newton(N)

2. Conditions 2(Second Operating conditions): 1650 rpm and 4200 N

3. Conditions 3(Third Operating conditions): 1500 rpm and 5000 N

There are two parts in this dataset: 1) Learning set, 2) Test set

Learning Set has the 6 bearing datasets directories: Bearing1\_1, Bearing1\_2, Bearing2\_1, Bearing2\_2, Bearing3\_1 and Bearing3\_2.

Test Set has the remaining 11 bearing datasets directories: Bearing1\_3, Bearing1\_4, Bearing1\_5, Bearing1\_6, Bearing1\_7, Bearing2\_3, Bearing2\_4,Bearing2\_5, Bearing2\_6, Bearing2\_7 and Bearing3\_3.

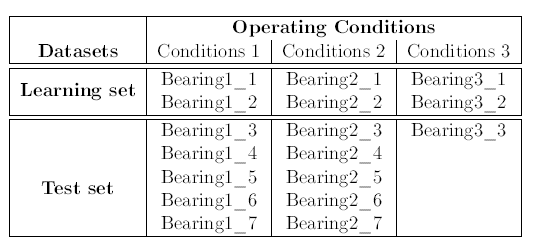


Table 1: Structure of PHM Dataset [9]

Notation: Here, Bearing1\_1 means Conditions 1 on the Bearing Number 1, Bearing 2\_1 means Conditions 2 on the Bearing Number 1 and so on.

Data Acquisition characteristics: Each of the bearing directory will have vibration ASCII files named “acc\_(4-digit number).csv” and temperature ASCII files named “temp\_(4-digit number).csv”.

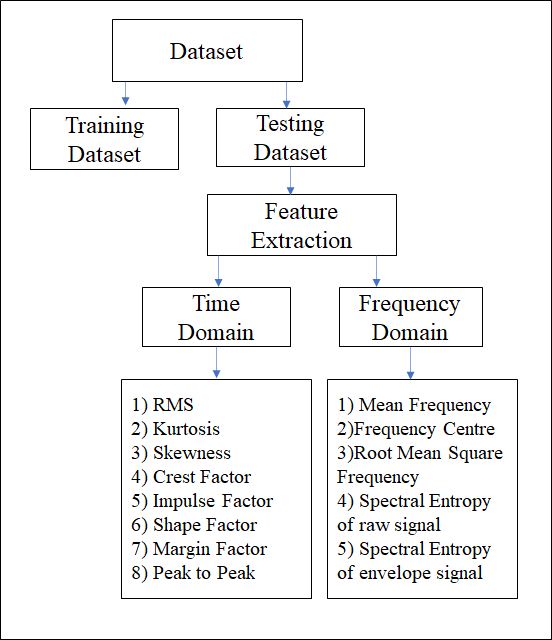
Vibration Signals (horizontal and vertical): Sampling frequency is 25.6 kHz and 2560 samples are recorded every 10 seconds.

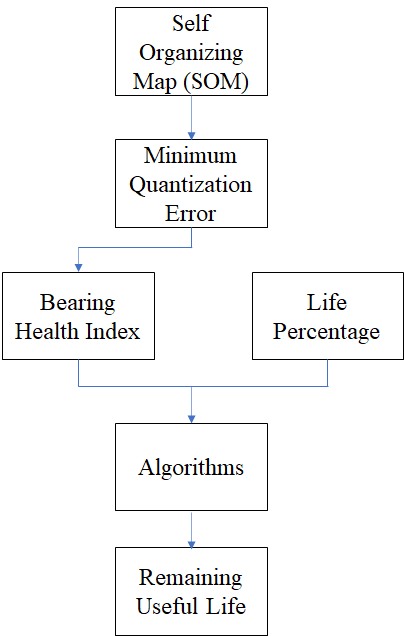
Details of each column in vibration signal and temperature signal data:

| Column | 1 | 2 | 3 | 4 | 5 | 6 |
| --- | --- | --- | --- | --- | --- | --- |
| Vibration Signal | Hour | Minute | Second | Microsecond | Horizontal Acceleration | Vertical Acceleration |

Table 2: Arrangement of data inside the ASCII csv file [9]

**Architecture**

****

****

**Approaches**

1. PHM RUL[6]: In this approach, we follow 3 methods:

1. Converting the PHM CSV data into the mat tables

2. Plotting all the bearing data using the mat tables

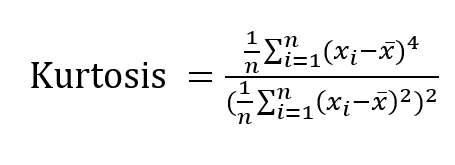
3. Remaining Useful Lifetime (RUL) prediction using Failure threshold and first predicting time

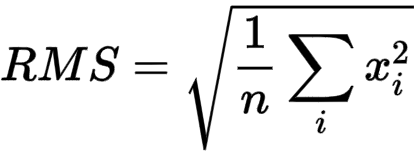
1. **Converting the PHM CSV data into the mat tables**
2. Here we consider the learning set and plot the vibration signals with respect to vertical acceleration and horizontal acceleration. Only acceleration data is considered for our experiment.
3. We concatenate all the horizontal and vertical signals from all the csv files from the Learning set and training set and plot them.
4. We convert the directories into the mat tables. e.g. Bearing1\_1.mat

**2) Plotting all the bearing data using the mat tables**

1. Plotting of horizontal and vertical signals is done on the mat tables of all the bearing directories.
2. Functions to clip the horizontal signals, calculating the kurtosis and root mean square (RMS) are carried out. Plotting of the same is done for the bearings belonging to conditions 1. i.e. Bearing1\_1, Bearing1\_2, Bearing1\_3 and Bearing1\_4.

Kurtosis is the measure of the “tailedness” of the probability distribution of a real-valued random variable. RMS is the square root of the mean square.





where n = number of measurements,=each value, x

Formulas:

1. sample\_kurtosis= sorted values of kurtosis from scipy.stats.kurtosis
2. sample\_rms= sorted values of rms
3. minimum\_kurtosis= mean of sample\_kurtosis - 3\* standard deviation of sample\_kutosis
4. maximum\_kurtosis= mean of sample\_kurtosis + 3\* standard deviation of sample\_kutosis
5. minimum\_rms= mean of sample\_rms - k\* standard deviation of sample\_rms
6. maximum\_rms= mean of sample\_rms + k\* standard deviation of sample\_rms

Plotting for horizontal signals, vertical signals, total absolute horizontal and vertical signals, clipped horizontal signals, total signals are carried out to understand the prediction pattern for all the bearings.

**3)** **RUL prediction using Failure threshold and first predicting time**

1. Converting the mat tables into arrays and calculating the first predicting time for the learning dataset
2. Calculating the maximum RUL = Failure Threshold - First Predicting Time,

predicted\_rms=checkpoint + predicted\_rul, true\_rms = checkpoint + true\_rul.

1. The following parameters are then plotted to get the results
2. RUL using Self Organizing Map (SOM) and regressors: The Self-Organizing Map and different Regression were used to learn and forecast the bearings' remaining useful life.

These steps were carried out to obtain the results

1. Feature Extraction: A time series of vibrational signals is used as the input dataset. As a result, it is necessary to extract characteristics from it in order to obtain relevant features that better characterise the data.The datasets have yielded a number of time-domain and frequency-domain features that have been extracted.

Time Domain features are: RMS, Kurtosis, Skewness, Crest Factor, Impulse Factor, Shape Factor, Margin Factor, Peak to Peak

Fourier Transforms, which translate data sets into constituent frequencies, were used to get the signal's frequencies. The Fourier Transform is a mathematical technique that allows you to break down a signal into its individual frequencies.

the frequency domain features are Mean frequency, Frequency centre, Root Mean Square frequency, Spectral Entropy of raw signal, Spectral Entropy of envelope signal

Each csv file corresponding to 10 second observations is converted to a feature vector consisting of features derived using the horizontal acceleration data in the csv file. As a result, all of the csv files produce an array of feature vectors.

2. Self Organizing Map(SOM):SOM is an adaptable unsupervised neural network that uses high-dimensional feature vectors to project them onto two-dimensional topological maps in order to classify input. The output layer of the SOM network is represented by topological maps, which are a two-dimensional array of neurons that are interconnected. A weight vector is associated with every neuron or node. The input feature space's dimension is the same as the weight vector's dimension. Kohonen's Map is the other name for it.

From the features extracted in the previous step, a set of healthy features is extracted that corresponds to the features that represent the bearing in a healthy state, i.e. it is not damaged. These are the first 1/5th feature vectors, as these will always be the features corresponding to the bearing in a healthy state. The Self-Organizing Map is trained using these healthy properties.

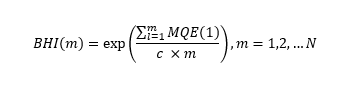
3. Minimum Quantization Error(MQE):The Euclidean distance between the input features and the matching BMU is the Minimum Quantization Error. The trained SOM is supplied with the remaining feature vectors (aside from healthy feature vectors) in order to calculate the BMI for each. For each feature vector, the appropriate MQE is calculated after obtaining the BMI. This creates a new feature that represents the departure of the inspected feature vector from the bearing's normal state; any change in the MQE value indicates the presence of a problem or an outlier.

4. Life Percentage: The life percentage represents the percentage of a bearing's life that has been completed at a given point in time. When a bearing is new, it has a life percentage of 0%, but when it is damaged or fails, it has a life percentage of 100%.

A bearing's life percentage is computed by dividing the present age of the bearing by the entire age of the bearing and then expressing the result as a percentage. The advantage of employing the life percentage is that no threshold is required for RUL estimate.

In our project, the life % is derived using the feature ‘minutes' that we calculated in each feature vector. The ‘minutes' feature in the training set's last feature vector corresponds to the features that appear after the bearing fails. As a result, we divide all of the "minutes" features by the "minutes" feature in the last feature, yielding the life percentage.

5. BHI (Bearing Health Index):The input vector for RUL estimation should be as monotonic as possible and rise gradually with the increase in bearing operating time. BHI is a monotonic index that corresponds to the bearing operation time. It describes a method for extracting monotonicity from MQE where BHI (m) is the health index value, MQE(l) is the MQE value at inspection point ‘l', BHI at the mth measurement point, N is the total number of measurement points, and BHI at the mth measurement point. A factor of c is used to adapt the BHI scale, which aids in preserving similar BHI trends for different bearings working under the same conditions and removing variances in BHIs in healthy bearing states. c is calculated as follows:



6. Different Algorithms:

1. Support vector regressor: As the name implies, Support Vector Regression is a regression algorithm that can handle both linear and non-linear regressions. This approach is based on the Support Vector Machine idea. SVR is different from SVM in that SVM is a classifier that predicts discrete categorical labels, whereas SVR is a regressor that predicts continuous ordered variables. The goal of basic regression is to reduce the error rate, however the goal of SVR is to fit the error within a particular threshold, which means that the goal of SVR is to approximate the best value within a given margin known as the - tube.
2. Random Forest Regressor: It is a supervised learning algorithm which uses ensemble learning methods to perform regression. It combines predictions from multiple machine learning techniques to make an accurate prediction. Steps include:
3. Selecting a random k data points from the training set.
4. Building the decision tree to these k data points.
5. Choose the number of trees N to build and repeat a and b.
6. For the new data point, make each of the data points of N-trees predict the value of y and assign the new data point to the average across all the predicted y values.
7. Decision Tree Regressor: Decision Tree is a decision-making tool which uses a flowchart structure in the form of decisions. It is also a supervised learning algorithm. The branches represent the result of the mode and the nodes will have conditions or results. A Decision Tree Regressor will observe the features of an object and train a model in the form of a tree to predict the outcomes of the future which will give continuous outputs.

4) Gradient Boosting Regressor: Gradient boosting Regression calculates the difference between the current prediction and the known correct target value. The difference is called residual. After that Gradient Boosting Regression trains a weak model that maps features to that residual. This residual predicted by a weak model is added to the existing model input and thus this process nudges the model towards the correct target. Repetition of this step improves the model's overall prediction.

5) Voting Regressor: A voting Regressor is an ensemble meta estimator that fits many base regressors, one on top of the other, on the entire dataset. The individual guesses are then averaged to give a final prediction.

7. Predicting RUL: First, the data from the test set is used to extract features. The corresponding BMU is detected using the trained SOM model once the feature vectors are obtained, and the related MQE values are calculated.

The BHI is calculated after the MQE values have been collected. The BHI value for the most recent observation is then fed into the trained SVR model, which outputs a normalised life percentage (life %/100).

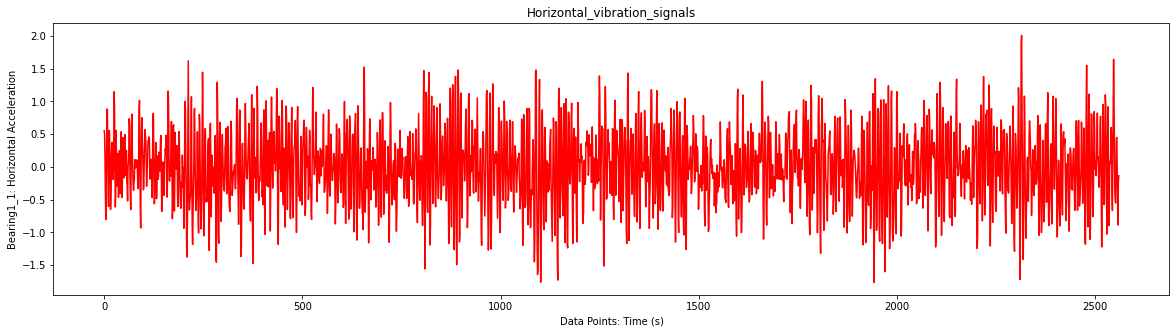
Once we obtain the life %, we already have the minute features, which correlate to the time when the observations were made and the characteristics were calculated. We may use these to calculate the minutes, or the time when the bearing will fail as predicted by our model's output, which equates to a 100% life percentage. So, in order to obtain the minutes, we will use the following procedure:

minutes at 100% = 100 \* (minutes at the last observations / predicted life percentage for last observation)

Once we have the minutes at 100%, which correspond to the minutes when the bearing is expected to fail, we can calculate the RUL by subtracting the minutes from the last observation from the minutes at 100%, which corresponds to the time left for the bearing to reach 100%.

**Results**

1. PHM RUL



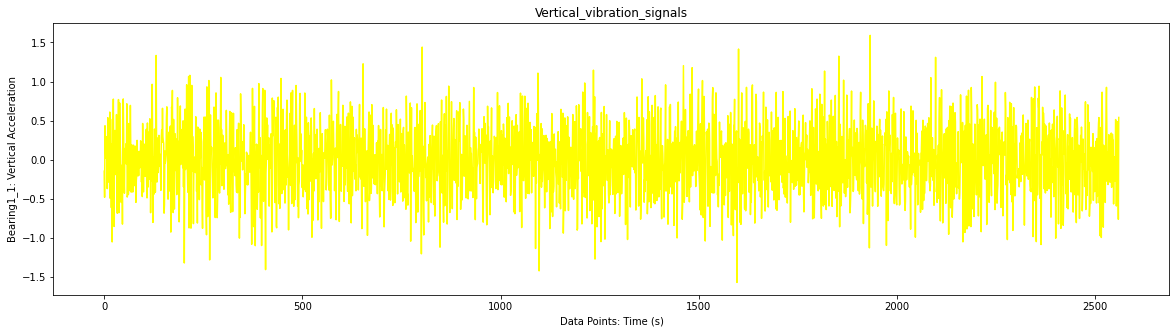
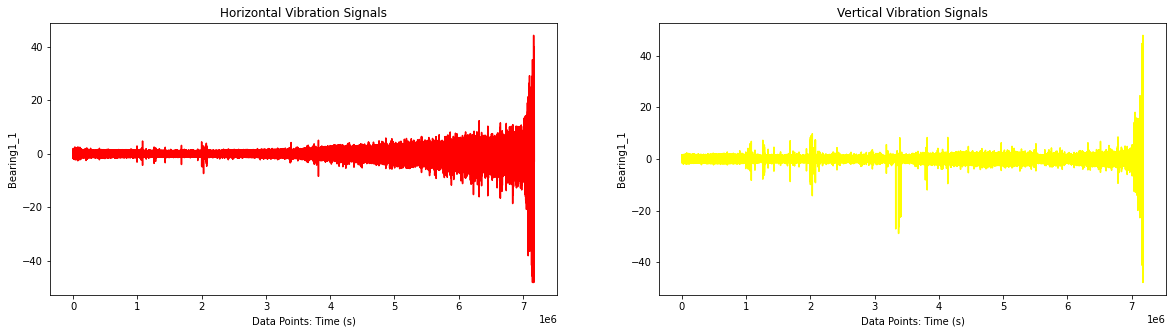


Fig 4:Horizontal and Vertical Acceleration for the Bearing1\_1 csv data



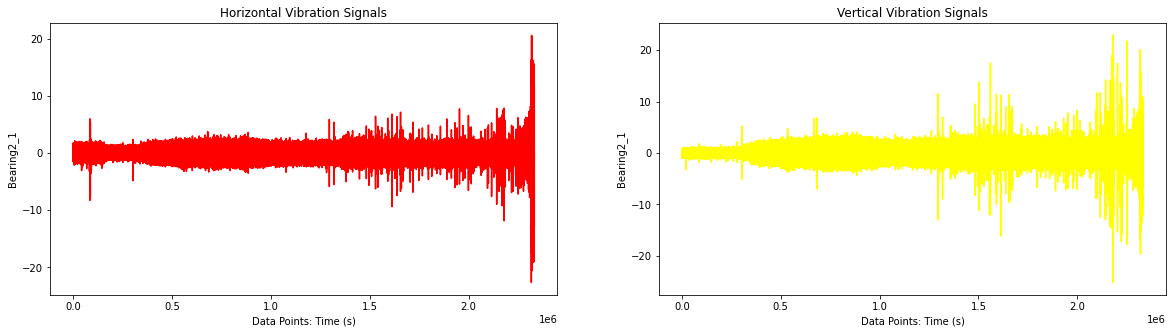
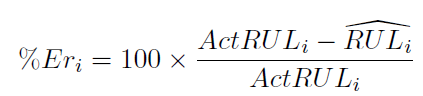
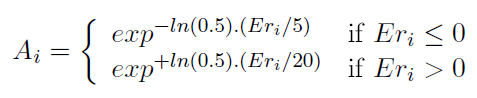


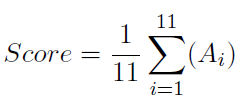
Fig 5: Plots of concatenated vibration signals of horizontal and vertical acceleration for Bearing 1\_1 and Bearing2\_1

2.RUL using Self Organizing Map (SOM) and regressors:

We had obtained different RUL results for 5 types of regressors used. We calculate percent error (%Eri), score of accuracy(Ai) and final score of RUL. Here Actual RUL to be estimated is the ground truth value

Percent error 

Score of Accuracy 

Final score of all RUL 

1. **Support Vector Regressor**

| Test Set | Actual RUL to be estimated in seconds | Predicted RUL (1st Model) in seconds | % Error (1st Model) | Ai (1st Model) | Predicted RUL (2nd Model) in seconds | % Error (2nd Model) | Ai (2nd Model) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Bearing1\_3 | 5730 | 2086 | 63.59511344 | 0.110356562 | 2536 | 55.7417103 | 0.144878406 |
| Bearing1\_4 | 339 | 1630 | -380.8259587 | 1.1803E-23 | 1300 | -283.480826 | 8.56552E-18 |
| Bearing1\_5 | 1610 | 4952 | -207.5776398 | 3.18113E-13 | 2571 | -59.68944099 | 0.000254881 |
| Bearing1\_6 | 1460 | 2852 | -95.34246575 | 1.81891E-06 | 3575 | -144.8630137 | 1.89836E-09 |
| Bearing1\_7 | 7570 | 4102 | 45.81241744 | 0.204387536 | 1957 | 74.14795244 | 0.076552982 |
| Bearing2\_3 | 7530 | 6892 | 8.472775564 | 0.745541839 | 2173 | 71.14209827 | 0.084958087 |
| Bearing2\_4 | 1390 | 909 | 34.60431655 | 0.301406863 | 1085 | 21.94244604 | 0.46744797 |
| Bearing2\_5 | 3090 | 10136 | -228.02589 | 1.86842E-14 | 2968 | 3.948220065 | 0.872114217 |
| Bearing2\_6 | 1290 | 1048 | 18.75968992 | 0.521961573 | 3305 | -156.2015504 | 3.94212E-10 |
| Bearing2\_7 | 580 | 1112 | -91.72413793 | 3.0037E-06 | 1575 | -171.5517241 | 4.69415E-11 |
| Bearing3\_3 | 820 | 839 | -2.317073171 | 0.72526763 | 2272 | -177.0731707 | 2.1834E-11 |
|  |  |  | Final Score | 0.237175166 |  |  | 0.14965514 |

Table 3: Results of Support Vector Regressor

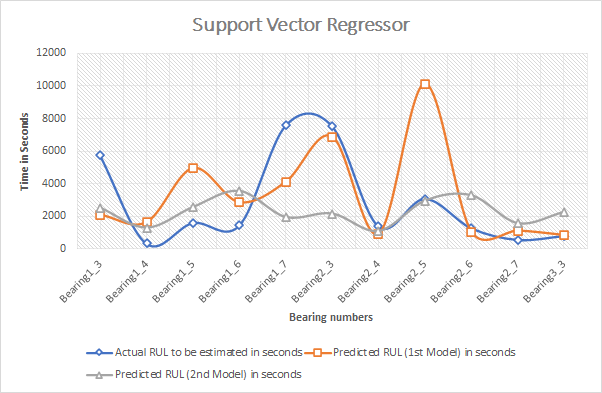


Fig 6: RUL prediction using SVR

By looking at the graphs of Actual RUL and the Predicted RUL, we observe that the support vector regressor of 1st model and 2nd model are doing good prediction for bearing number 1\_4, 1\_6, 2\_4, 2\_6, 2\_7 and 3\_3 but for the bearing number 1\_3, 1\_5, 1\_7, 2\_3 and 2\_5 the models are showing less accuracy in predicting RUL.

1. **Random Forest Regressor**

| Test Set | Actual RUL to be estimated in seconds | Predicted RUL (1st Model) in seconds | % Error (1st Model) | Ai (1st Model) | Predicted RUL (2nd Model) in seconds | % Error (2nd Model) | Ai (2nd Model) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Bearing1\_3 | 5730 | 495 | 91.36125654 | 0.042157618 | 0 | 100 | 0.03125 |
| Bearing1\_4 | 339 | 776 | -128.9085546 | 1.73353E-08 | 0 | 100 | 0.03125 |
| Bearing1\_5 | 1610 | 2966 | -84.22360248 | 8.49638E-06 | 0 | 100 | 0.03125 |
| Bearing1\_6 | 1460 | 993 | 31.98630137 | 0.330033627 | 0 | 100 | 0.03125 |
| Bearing1\_7 | 7570 | 2506 | 66.89564069 | 0.098428367 | 742 | 90.19815059 | 0.043891716 |
| Bearing2\_3 | 7530 | 0 | 100 | 0.03125 | 15 | 99.80079681 | 0.031466492 |
| Bearing2\_4 | 1390 | 0 | 100 | 0.03125 | 561 | 59.64028777 | 0.126568089 |
| Bearing2\_5 | 3090 | 0 | 100 | 0.03125 | 25 | 99.19093851 | 0.032138649 |
| Bearing2\_6 | 1290 | 565 | 56.20155039 | 0.142587803 | 2086 | -61.70542636 | 0.000192736 |
| Bearing2\_7 | 580 | 721 | -24.31034483 | 0.034385188 | 1301 | -124.3103448 | 3.27923E-08 |
| Bearing3\_3 | 820 | 472 | 42.43902439 | 0.229735989 | 1461 | -78.17073171 | 1.96631E-05 |
|  |  |  | Final Score | 0.088280646 |  |  | 0.03266158 |

Table 4: Results of Random Forest Regressor

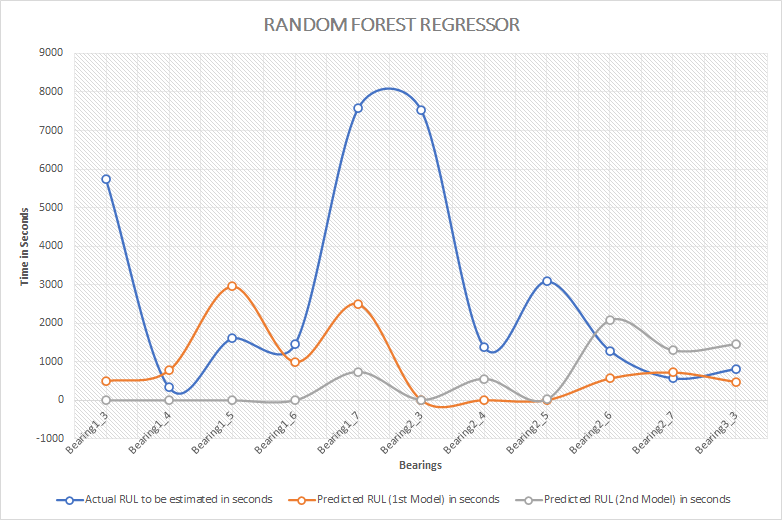


Fig 7: RUL prediction using Random Forest Regressor

By looking at the graphs of Actual RUL and the Predicted RUL, we observe that the support vector regressor of 1st model and 2nd model are doing good prediction for bearing number 1\_4, 1\_6, 2\_4, 2\_6, 2\_7 and 3\_3 but for the bearing number 1\_3, 1\_5, 1\_7, 2\_3 and 2\_5 the models are showing less accuracy in predicting RUL.

1. **Decision Tree Regressor**

| Test Set | Actual RUL to be estimated in seconds | Predicted RUL (1st Model) in seconds | % Error (1st Model) | Ai (1st Model) | Predicted RUL (2nd Model) in seconds | % Error (2nd Model) | Ai (2nd Model) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Bearing1\_3 | 5730 | 495 | 91.36125654 | 0.042157618 | 0 | 100 | 0.03125 |
| Bearing1\_4 | 339 | 776 | -128.9085546 | 1.73353E-08 | 0 | 100 | 0.03125 |
| Bearing1\_5 | 1610 | 2966 | -84.22360248 | 8.49638E-06 | 0 | 100 | 0.03125 |
| Bearing1\_6 | 1460 | 993 | 31.98630137 | 0.330033627 | 0 | 100 | 0.03125 |
| Bearing1\_7 | 7570 | 2506 | 66.89564069 | 0.098428367 | 742 | 90.19815059 | 0.043891716 |
| Bearing2\_3 | 7530 | 0 | 100 | 0.03125 | 15 | 99.80079681 | 0.031466492 |
| Bearing2\_4 | 1390 | 0 | 100 | 0.03125 | 561 | 59.64028777 | 0.126568089 |
| Bearing2\_5 | 3090 | 0 | 100 | 0.03125 | 25 | 99.19093851 | 0.032138649 |
| Bearing2\_6 | 1290 | 565 | 56.20155039 | 0.142587803 | 2086 | -61.70542636 | 0.000192736 |
| Bearing2\_7 | 580 | 721 | -24.31034483 | 0.034385188 | 1301 | -124.3103448 | 3.27923E-08 |
| Bearing3\_3 | 820 | 472 | 42.43902439 | 0.229735989 | 1461 | -78.17073171 | 1.96631E-05 |
|  |  |  | Final Score | 0.088280646 |  |  | 0.03266158 |

Table5 : Results of Decision Tree Regressor

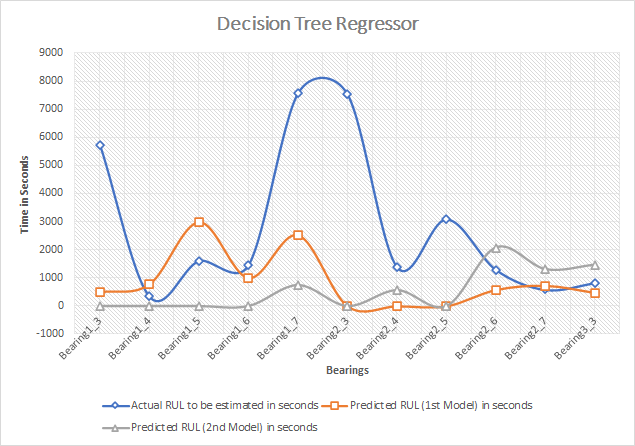


Fig 8: RUL prediction using Decision Tree Regressor

By looking at the graphs of Actual RUL and the Predicted RUL, we observe that the support vector regressor of 1st model and 2nd model are doing good prediction for bearing number 1\_4, 1\_6, 2\_4, 2\_6, 2\_7 and 3\_3 but for the bearing number 1\_3, 1\_5, 1\_7, 2\_3 and 2\_5 the models are showing less accuracy in predicting RUL.

1. **Gradient Boosting Regressor**

| Test Set | Actual RUL to be estimated in seconds | Predicted RUL (1st Model) in seconds | % Error (1st Model) | Ai (1st Model) | Predicted RUL (2nd Model) in seconds | % Error (2nd Model) | Ai (2nd Model) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Bearing1\_3 | 5730 | 10385 | -81.2390925 | 1.28505E-05 | 10349 | -80.61082024 | 1.40199E-05 |
| Bearing1\_4 | 339 | 6562 | -1835.693215 | 3.0217E-111 | 6539 | -1828.908555 | 7.7399E-111 |
| Bearing1\_5 | 1610 | 13859 | -760.8074534 | 1.56613E-46 | 13222 | -721.242236 | 3.77477E-44 |
| Bearing1\_6 | 1460 | 13268 | -808.7671233 | 2.0294E-49 | 13222 | -805.6164384 | 3.14092E-49 |
| Bearing1\_7 | 7570 | 9041 | -19.4319683 | 0.067620587 | 8625 | -13.93659181 | 0.144855025 |
| Bearing2\_3 | 7530 | 6901 | 8.353253652 | 0.748636512 | 6905 | 8.300132802 | 0.750016043 |
| Bearing2\_4 | 1390 | 3511 | -152.5899281 | 6.50386E-10 | 3513 | -152.7338129 | 6.37541E-10 |
| Bearing2\_5 | 3090 | 11499 | -272.1359223 | 4.12843E-17 | 11505 | -272.3300971 | 4.01878E-17 |
| Bearing2\_6 | 1290 | 3281 | -154.3410853 | 5.102E-10 | 3576 | -177.2093023 | 2.14258E-11 |
| Bearing2\_7 | 58 | 1070 | -1744.827586 | 8.9312E-106 | 1164 | -1906.896552 | 1.5609E-115 |
| Bearing3\_3 | 820 | 2103 | -156.4634146 | 3.80158E-10 | 2197 | -167.9268293 | 7.75885E-11 |
|  |  |  | Final Score | 0.074206359 |  | -556.2828274 | 0.08135319 |

Table 6: Results of Gradient Boosting Regressor

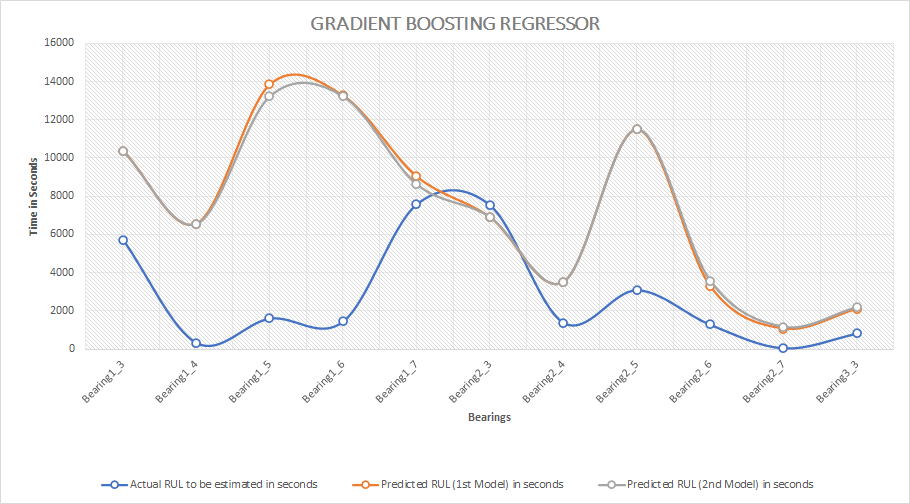


Fig 9: RUL prediction using Gradient Boosting Regressor

By looking at the graphs of Actual RUL and the Predicted RUL, we observe that the support vector regressor of 1st model and 2nd model are doing good prediction for bearing number 1\_7 and 2\_3 but for the bearing number 1\_3, 1\_4, 1\_5, 1\_6, 2\_4, 2\_5, 2\_6, 2\_7 and 3\_3 the models are showing less accuracy in predicting RUL.

1. **Voting Regressor**

| Test Set | Actual RUL to be estimated in seconds | Predicted RUL (1st Model) in seconds | % Error (1st Model) | Ai (1st Model) | Predicted RUL (2nd Model) in seconds | % Error (2nd Model) | Ai (2nd Model) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Bearing1\_3 | 5730 | 2709 | 52.72251309 | 0.160859645 | 2504 | 56.30017452 | 0.142101262 |
| Bearing1\_4 | 339 | 2082 | -514.159292 | 1.10797E-31 | 1506 | -344.2477876 | 1.88026E-21 |
| Bearing1\_5 | 1610 | 5586 | -246.9565217 | 1.35437E-15 | 3030 | -88.19875776 | 4.89672E-06 |
| Bearing1\_6 | 1460 | 3736 | -155.890411 | 4.11587E-10 | 3282 | -124.7945205 | 3.06635E-08 |
| Bearing1\_7 | 7570 | 4213 | 44.34610304 | 0.215042668 | 2517 | 66.75033025 | 0.09892531 |
| Bearing2\_3 | 7530 | 2679 | 64.42231076 | 0.107237728 | 1792 | 76.20185923 | 0.071293138 |
| Bearing2\_4 | 1390 | 862 | 37.98561151 | 0.268077013 | 1261 | 9.28057554 | 0.724958956 |
| Bearing2\_5 | 3090 | 4253 | -37.63754045 | 0.005419937 | 2829 | 8.446601942 | 0.746218433 |
| Bearing2\_6 | 1290 | 1214 | 5.891472868 | 0.815313245 | 2704 | -109.6124031 | 2.5158E-07 |
| Bearing2\_7 | 580 | 893 | -53.96551724 | 0.000563576 | 1328 | -128.9655172 | 1.7199E-08 |
| Bearing3\_3 | 820 | 884 | -7.804878049 | 0.338921811 | 1820 | -121.9512195 | 4.54784E-08 |
|  |  |  | Final Score | 0.173766875 |  | Final Score | 0.162136576 |

Table 7: Results of Voting Regressor

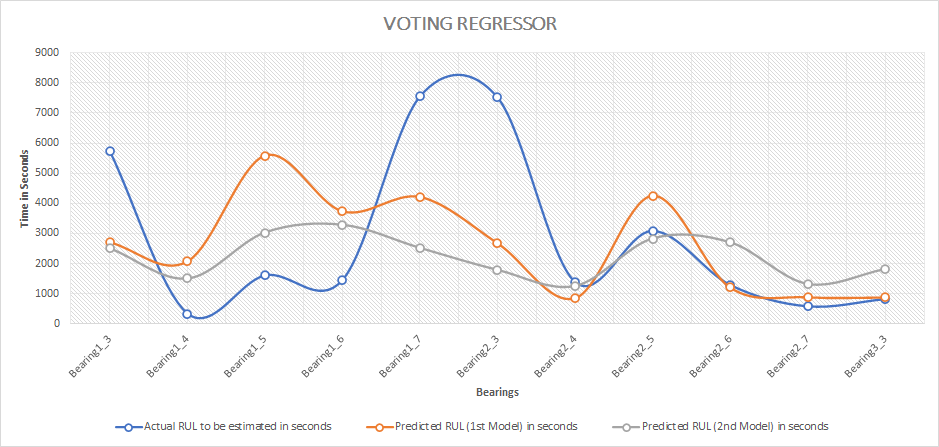


Fig 10: RUL prediction using Voting Regressor

By looking at the graphs of Actual RUL and the Predicted RUL, we observe that the support vector regressor of 1st model and 2nd model are doing good prediction for bearing number 2\_4 and 2\_7 but for the bearing number 1\_3, 1\_4, 1\_5, 1\_6, 1\_7, 2\_3, 2\_5, 2\_6, and 3\_3 the models are showing less accuracy in predicting RUL.

**Conclusion**

The predictive maintenance model gives more accurate results for Support Vector Regressor (SVR), Random Forest Regressor and Decision Tree Regressor algorithms for predicting Remaining Useful Lifetime (RUL) for 6 bearings. The final score of accuracy was highest for the Support Vector Regressor (SVR) algorithm for the 1st and 2nd model. Also the final score was better in case of Voting Regressor.

**Future Scope**

This method can be implemented on other regressors like AdaBoostRegressor, Light Gradient Boosting Machine (LGBM) Regressor, XGBM Regressor etc. This technique can be implemented on other instruments like gears, shafts, and other machines apart from the bearings.

**References**

1. Susilo, D. D., Widodo, A., Prahasto, T., & Nizam, M. (2017). Fault diagnosis of roller bearing using parameter evaluation technique and multi-class support vector machine. doi:10.1063/1.4968334
2. Kamiel, B. P., & Howard, I. (2019). Ball bearing fault diagnosis using wavelet transform and principal component analysis. THE 4TH BIOMEDICAL ENGINEERING’S RECENT PROGRESS IN BIOMATERIALS, DRUGS DEVELOPMENT, HEALTH, AND MEDICAL DEVICES: Proceedings of the International Symposium of Biomedical Engineering (ISBE) 2019. doi:10.1063/1.5138361
3. Glowacz, A. (2019). Acoustic fault analysis of three commutator motors. Mechanical Systems and Signal Processing, 133, 106226. doi:10.1016/j.ymssp.2019.07.007
4. Cheng, H., Zhang, Y., Lu, W., & Yang, Z. (2019). A bearing fault diagnosis method based on VMD-SVD and Fuzzy clustering. International Journal of Pattern Recognition and Artificial Intelligence. doi:10.1142/s0218001419500186
5. Elhaj, M., Gu, F., Ball, A. D., Albarbar, A., Al-Qattan, M., & Naid, A. (2008). Numerical simulation and experimental study of a two-stage reciprocating compressor for condition monitoring. Mechanical Systems and Signal Processing, 22(2), 374–389. doi:10.1016/j.ymssp.2007.08.003
6. [GitHub - famer3riots/PHM-RUL-Prediction-by-EM: Remain useful life prediction of rolling bearing.](https://github.com/famer3riots/PHM-RUL-Prediction-by-EM)
7. FEMTO Bearing Dataset: [Prognostics Center of Excellence - Data Repository (nasa.gov)](https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#bearing)
8. [GitHub - linzhang92/phm-ieee-2012-data-challenge-dataset: Dataset that was used during the PHM IEEE 2012 Data Challenge, built by the FEMTO-ST Institute](https://github.com/linzhang92/phm-ieee-2012-data-challenge-dataset)
9. Patrick Nectoux, Rafael Gouriveau, Kamal Medjaher, Emmanuel Ramasso, Brigitte Morello, Noureddine Zerhouni, Christophe Varnier. PRONOSTIA: An Experimental Platform for Bearings Accelerated Life Test. IEEE International Conference on Prognostics and Health Management, Denver, CO, USA, 2012