7th International Conference on Reliability and Safety Engineering (ICRSE2023)

Remaining Useful Life Prediction of Industrial Rolling Bearings with Limited Offline History using Stacking Regressor

Ali Yavari [1], Mehdi Behzad [1*], Hesam Addin Arghand [2], Somaye Mohammadi [1], Vahid Gholami[3]

1- Condition Monitoring Lab, Mechanical Engineering Department, Sharif University of Technology, Tehran, Iran
2- Mechanical Engineering Department, Faculty of Engineering, University of Zanjan, Zanjan, Iran
3- Imam Khomeini Oil Refining Company of Shazand, Arak, Iran

Corresponding author, m_behzad@sharif.edu

Abstract

Predicting the remaining useful life (RUL) of ball bearings to take correct and timely actions plays a vital role in the reliability of industrial rotating equipment. Artificial intelligence and machine learning algorithms' ability to interpret vibration data has made researchers employ them for RUL prediction. Most of the research activities in the intelligent monitoring of rolling element bearings have been developed to interpret online data with short time intervals and often for laboratory accelerated life data. However, industrial rotating equipment is often monitored offline with periodic measurements and relatively long time intervals, due to the high cost of online condition monitoring compared to offline condition monitoring. In the offline method, the number of measurements that are obtained during the deterioration of the rolling element bearing is much more limited than in the online mode, which makes predictions difficult. To cope with this problem, a novel stacked model algorithm is proposed here and is trained and tested using real industrial data. The proposed model contains three base models: relevance vector regressor, regression-enhanced XGBoost, linear Ridge, and a linear Ridge regressor meta-model. The results show that the proposed technique can improve the RUL prediction accuracy to 89% compared to simple conventional algorithms. The results of this algorithm are underestimated and give the necessary warning before catastrophic failures. Also, the proposed model can provide better generalization by avoiding overfitting.

Keywords: Offline monitoring; Remaining useful life; Rolling bearings; Stacking regressor; Vibration analysis

Introduction

Condition-based maintenance (CBM) is a strategy that monitors the health status of machines over time and makes optimal decisions based on it. The effectiveness of this technique in reducing unnecessary repairs and increasing the reliability of machines has made it popular in recent years. Considering the need of the industries to increase the reliability and availability of machines, they have decided to move towards newer and more reliable techniques. In the meantime, CBM can establish its position in the industry and plays an effective role in reducing repair costs.

Rolling bearings are among the most common mechanical parts in rotating machines. Their failure is the main factor in 45 to 55 percent of the failure of these machines [1]. Unforeseen failure can lead to a reduction in industrial machines' life and impose heavy costs due to non-production in industrial units. Therefore, estimating the remaining useful life of rolling bearings plays a vital role in the availability of rotating equipment in industrial factories. In this regard, the emergence of artificial intelligence algorithms and their widespread condition monitoring has led to a huge transformation in this field. Most of these algorithms function correctly just in case they are provided with a relatively large amount of data, like most deep learning algorithms.

During the last two decades, many algorithms have been developed in the field of intelligent monitoring of rolling bearings. Among these algorithms, some of them are more popular among researchers, like support vector machines (SVM). Isometric dimension reduction (ISOMAP) and support vector regression (SVR) methods have been used for predicting the RUL of rolling element bearings [2]. Multilayer perceptron (MLP) has been utilized to estimate the RUL of the timing belt in an internal combustion engine [3]. In addition to conventional machine learning methods, advanced deep learning techniques have also been employed for life prediction. These algorithms are usually used when facing a large data structure for the interpretation of online data - measurements with short time intervals and often on laboratory accelerated life data [4].

There are difficulties in acquiring industrial data in the field of condition monitoring. First, the deterioration process of machinery is a long-term phenomenon and takes months or even years. Thus, storing all the process data is time-consuming and costly. Secondly, machines are not allowed to go through a complete deterioration process because any unforeseen failure can lead to a complete stoppage of the production line or even catastrophic incidents. Third, many data are collected when the machine is out of service, which differs from the data under service. Finally, few organizations are willing to release their data for academic purposes, resulting in few data sources available to researchers.[5] So, researchers are faced with offline data - periodic measurements with relatively long time intervals e.g. monthly. This issue has pushed some researchers to develop techniques for predicting RUL based on the limited number of data.

In one of the first studies to predict the remaining life of rolling element bearings using limited data, the Weibull failure rate function is used in developing the RUL predictor [6]. The generalized Weibull function has also been used to create several auxiliary data to train the model using feed-forward neural networks (FFNN) [7]. In other research [8], two main uncertainties caused by measurement and process noise have been considered to overcome the shortcoming of the FFNN in predicting the condition of rolling elements in the presence of offline data collection.

The limited number of data available in offline monitoring which is the common method of CBM in most industrial plants will reduce the efficiency of the proposed algorithms significantly. Additionally, most of these algorithms e.g. decision tree-based algorithms cannot extrapolate out of the training dataset. They should be trained with a complete run-to-failure history which is not available in most industrial plants because vital machines are not allowed through a complete degradation process. Therefore, this research aims to provide a solution to be able to work on a limited number of vibration signals as well as the ability to predict out of the training set. This is a novel problem in the field of offline condition monitoring which this research is going to provide the answer to.

This paper is organized as follows: The next section introduces the datasets used to train the RUL prediction model. Section 3 presents the features selected to be used as input to the model. Section 4 focuses on the model structure for RUL prediction. Section 5 discusses the prediction results and evaluates them from different perspectives. In the end, a summary, conclusion, future work, and related references are given.

Bearing Run-to-Failure Case

The industrial dataset used in this paper has been collected at the Imam Khomeini oil refining company of Shazand, including 19 run-to-failure data of rolling element bearings. The period between each data collection is nearly two weeks with a sampling frequency of 20 kHz. After predicting bearing failures, they have been disassembled and the types of defects related to them have been verified.

This paper is intended to predict the RUL based on the history of the corresponding bearing. So, the performance of the proposed model is shown only for one bearing data. Among the available industrial data, a bearing is selected with the frequency spectrum shown in **Fig. 1,** extracted from the VibroDesigner software. The measurements have been carried out in two vertical and horizontal directions (above and bottom graphs). The black and red signals in each graph correspond to the vibration acceleration and velocity, respectively. Acceleration signals are usually used in life prediction matters due to the high-frequency content of bearing vibrations. In this research, only the vertical acceleration signal has been used to predict the RUL based on its history, since both signals show approximately the same trend.

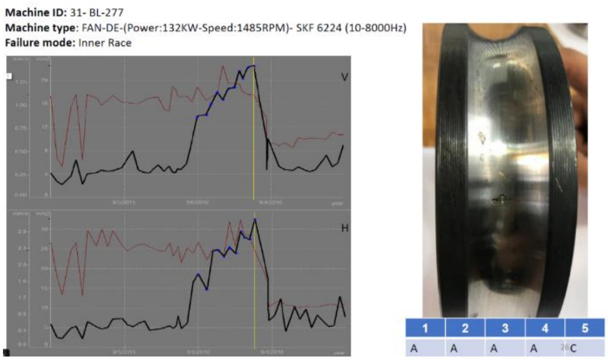

Fig. 1. Industrial bearing with its vibration trends in two directions, selected to be used in RUL prediction- black graph: acceleration, red graph: velocity.

The selected bearing is related to the drive end (DE) point of a fan whose power consumption is 132 kilowatts and its speed is 1485 rpm. As it is clear from the figure, the failure mode of this bearing is a defect of its inner ace.

Feature Selection

The first thing to consider while constructing a machine learning prediction model is what the input features of the model are. A failure threshold should be specified for each selected feature. This fact brings about serious issues in increasing the number of our input features. Many research studies have only used RMS as a single feature to construct their models. Because according to rotating machinery standards, such as ISO 10816-3, or the manufacturer's guidelines, RMS seems an appropriate feature to define the failure thresholds of machines.

When only a feature is used instead of the whole signal, a large part of the signal information is lost, which increases the prediction error. This issue does not have a strong impact when sufficient bearing history is available, as the measurement period is short. But facing a severe lack of data, using only RMS causes a large. One of the best features that model the degradation process well is the first principal component (PC_1) of signal features. Nevertheless, Determining the appropriate threshold level for PC_1 is a laborious task with many challenges in practice.

Pearson's correlation coefficient (PCC) and Copula models have been used to calculate a joint probability distribution function and analyze the correlation between PC_1 and time characteristics [9]. Having the joint probability distribution function, by assuming a value for one of the parameters, the conditional probability distribution function of the other variable can be obtained. The feature with the highest correlation can be used instead of PC_1. This research illustrated that PEAK has the most correlation with PC_1.

According to the mentioned noted, RMS and Peak of the acceleration signals are selected as the model input to predict the rolling bearing.

Proposing an RUL Predictor

An appropriate algorithm can predict the bearing life with sufficient reliability and accuracy. The algorithm is intended to be trained and work properly with a small number of offline data. This research proposes stacking regressors for this aim. Models that are built based on the linear combination of several prediction models intending to increase the performance are called stacking regressors. In these models, the goal is to maximize the accuracy of predictions by using several machine learning algorithms which are assigned weights based on their performance on the training set. Some metrics are needed to measure the performance. In this work, Root Mean Square error (RMSE) and coefficient of determination (R^2) are used to measue each model's performance.

There are two kind of predicting algorithms in a stacking model. First, "Base models" which are trained using training data and then are used to produce initial predctions. Second, the "Meta model" which is learned to fuse the predictions of "Base Models" in the best way possible by assigning the weights to the prediction of each model based on their performance on training set and then provide the final predictions. Fig. 2. shows the described procedure as well as the structure of a stacked model.

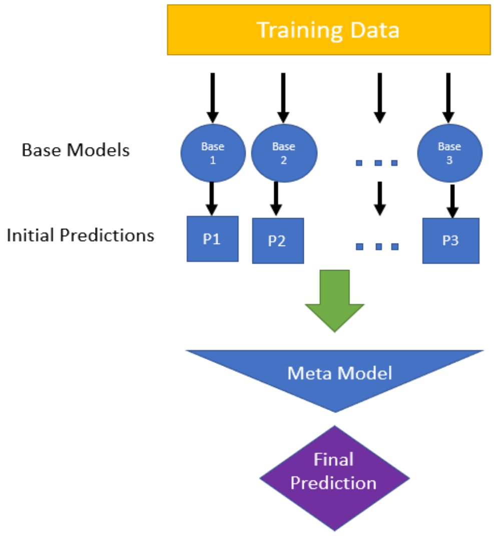

Fig. 2. General structure of the stacked model

The impotant thing that should be taken into consideration is that stacking models produce better results than a single algorithm when using uncorrelated or low-correlated base models. In this respect, model selection is an important criterion in stacked models. In practice, this is done by trial and error. This paper proposes a novel stacked model based on three base models which incluedes 1- Linear Ridge, 2- Regression Enhanced XGBoost (REXGB), 3-Relevance Vector Regression (RVR). In what follows each base models are explained.

1) Linear Ridge : One of the ways to prevent overfitting is to use regulating expressions in the cost function of the model.

$$E(w) = \sum_{i=1}^{N} (\hat{y} - w^T X)^2 + \lambda \sum_{j=0}^{M-1} |\omega_j|^q \quad (1)$$

where E, w, $\hat{\boldsymbol{y}}$, X, and $\boldsymbol{\lambda}$, respectively, are cost function, unknown model's weights, estimated RUL, input features, and regularization factor. The parameter q can be set to an arbitrary value. When q is equal to 1, it is called L1 or Lasso Regression, and when q is 2, it is called L2 or Ridge Regression.

2) Regression Enhanced Models : The concept of enhanced regression models has been first introduced by Hoazhe Zhang in [10]. This algorithm combines linear algorithms and ensemble learning algorithms based on regression trees e.g. random forests and gradient boosted algorithm. Decision tree-based algorithms, called non-parametric algorithms do not follow a specific mathematical function and do not make any assumptions about the data distribution. These characteristics along with the advantage of the high flexibility of the algorithm for modeling the data non-linearities well. However, the disadvantage is the lack of prediction outside the scope of training, which makes it impossible to predict the RUL of ballbearings with no historical data. One proposed method to cope with this problem is combining the algorithms with parametric models including penalties.

After some trial and error, the combination of the XGboost algorithm, an ensemble learning algorithm based on regression trees, and Linear Lasso, is proposed as one of the base models. The combination of the two mentioned algorithms results in an algorithm named REXGB (Regression Enhanced XGBoost) with detailed information in [10].

3) RVR (Relevance Vector Regression) : SVMs are among the most famous conventional machine learning algorithms which have been widely used in the field of RUL prediction. Bayesian formulation of Relevance vector machines (RVM) provides a sparser solution than SVM with fewer hyperparameters. RVR (relevance vector regressor) performs better with a limited number of data than SVR. Based on the mentioned points and given that RVR is a nonlinear parametric algorithm, it seems a good candidate as the third base model.

To conclude, in stacking regressor the aim is to increase the performance and decresing the varience of results by stacking different machine learning algorithms. Decreasing the variance is equal to more robust results by which critical decisions can be made. In the simplest form stacking regressor can be seen as getting the average from prediction of base models. It can be shown that the Variance of average is be bounded as shown in Eq. 2

$$Var(\bar{X}) \leq \frac{\sigma^2}{N} + \frac{N-1}{N}\sigma^2\rho \qquad (2)$$

Where $\boldsymbol{\sigma^2}$ is the variance of random variables participated in average, N is the number of available data, and $\boldsymbol{\rho}$ is the correlation factor. It is obvious that as correlation is getting near zero the variance of average decreses. This is ture for stacking regressors. In this regard three base models are used with low correlation. Liner Ridge is a linear parametric algorithm, RVR is a nolinear parametric algorithms, and REXGB is a nonparametric algorithm which can be suitable models with low corrlantion due to the fundamental difference in their training procedure.

After determining the base models it comes to determine a meta model. Selecting a meta-model is more of an experimental work. In most articles concerning stack models, the meta-model is a simple model such as linear regression for regression applications and logistic regression for classification tasks. Since the base models are all trying to predict the same relationship, the correlation between them is expected. Linear regression may make the final prediction too sensitive to data changes and poor generalization. In contrast, Ridge regression associated with regularization parameters can deal with the correlation between the predictions of each base model much better than linear regression. So, Ridge regression is selected as a meta-model in this research. Fig. 3 illustrates the structure of the proposed stacked model with three base models including REXGB, RVR, and linear Ridge.

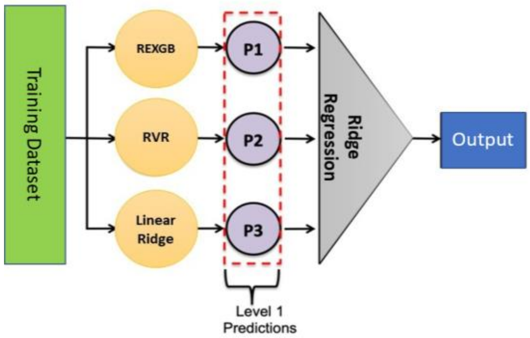

Fig. 3. Proposed stacked model for RUL prediction

Data Preparation

After constructing the prediction model, the model should be trained using the given vibration signals. By performing the initial signal processing techniques, signal features are extracted to be used in the process of training. The RUL prediction is performed in the zone where the degradation process has started. In this case, after appearing the first symptoms of degradation, the point after that is included in process of prediction. Fig.4 shows the health division along with the training and test points. The last five points of measurement are used for testing the model which is nearly seventy days before occurring the failure.

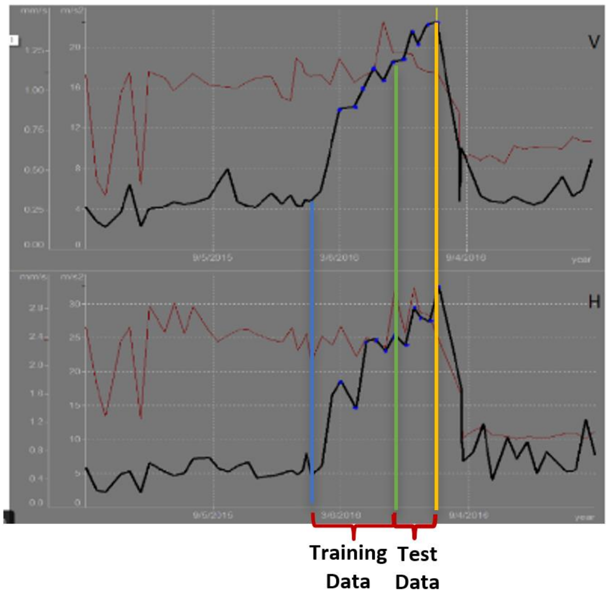

Fig. 4. Heatlth division of bearing vibration trend

RUL Prediction Results and Discussion

The failure thresholds of 23 m/s^2 for RMS and 134 m/s^2 for the peak of the vibrational acceleration of the bearing. These failure thresholds are obtained through the suggestion of the experts and the degradation trend of the ball bearings.

For evaluating the efficiency of the proposed method, each of the base algorithms is used for RUL prediction separately and then their results are compared to the result of the Stacking Regressor. First, the hyperparameters of each model are tuned using the grid search optimization technique. **Table. 1** to **Table. 3** show the optimal hyperparameters of each algorithm.

Table 1. Optimal parameters of the RVR algorithm

kernel	‘Sigmoid’
Coef	0.01
Gamma	0.01

Table 2. Optimal parameters of the REXGB algorithm

alpha	0.01
Objective	'reg:squarederror'
max_depth	10
n_estimators	1000
learning_rate	0.01

Table 3. Optimal parameters of Ridge algorithm

alpha	0.1

Two additional parameters should be optimized for stacking regressors comprising the number of cross-folds used in process of cross-validation and the regularization factor used in the meta-model. It is worth mentioning that the regularization term in the meta-model is usually a high number when there is a huge correlation between the prediction of base models. **Table. 4** gives the optimal value of the mentioned parameters.

Table. 4. Optimal values of the number of cross-folds and regularization terms in the meta-model

CV	8
alpha	500

The coefficient of determination, R2, which states the overall quality of regression and root mean squared error (RMSE) are first used as criteria for evaluating the results. **Table. 5** and **Table. 6** present the comparison of the models based on the criteria. The value of R2 higher than 0.9 shows that the regression model has an acceptable quality in describing the relationship between the inputs and output. So, the proposed model is acceptable in this view. Table 6 depicts that the proposed model provides approximately the same results for the train and test data with higher accuracy compared to the single models that are desired.

Table. 5. Comparison of models based on determination coefficient

Model	R2
RVR	0.87
REXGB	0.99
Linear Ridge	0.87
The proposed model (stacking regressor)	0.93

The RMSE on the test set for stacking regressor is something around 15 days. It means that the prediction can be accompanied by 15 days error. It must be noted that since the number of training data is so limited, higher accuracy in the results cannot be expected. The other point to be noticed is that the period between the two measurements is nearly two weeks (14 days) and the failure time is predicted around 70 days before it occurs. So, the accuracy of the developed model is acceptable and it can be employed in decisions regarding the working condition of the machine.

Table. 6. Comparison of models based on RMSE

Model	RMSE (RUL days)	
	Train data	Test data
RVR	20.7	33.9
REXGB	0.13	16.8
Linear Ridge	20.6	29.5
The proposed model (stacking regressor)	15.5	15.6

One of the major challenges in machine learning is overfitting which occurs when a model has a good performance on the training data but performs not so well on the test results. Most algorithms suffer from this phenomenon since the RMSE on the training set is meaningfully less than the RMSE on the test set. In the case of the stacking regressor that is proposed in this paper, the model has similar performance on both test and training data. Therefore, it can be concluded that if the base models are chosen carefully and their hyperparameters are optimized properly, it can result in a more efficient model. So, by implementing the stacking correctly, the prediction variance can be decreased and in turn, the reliability of the outcomes will increase.

Fig. 5 shows the predicted RUL vs ground truth RUL. The red line is the zone that predicted RUL equals the ground truth RUL. So, minimizing the distance of the points to this line is desired. The RMSE calculates these distances. Regardless of the point that the point is placed above or below the line. In practice, it is critically important to be under the line, meaning underestimating the prediction is underestimated or the model predicts the failure before it occurs and causes catastrophic failures. It can be seen in the figure that the model is desirably underestimated.

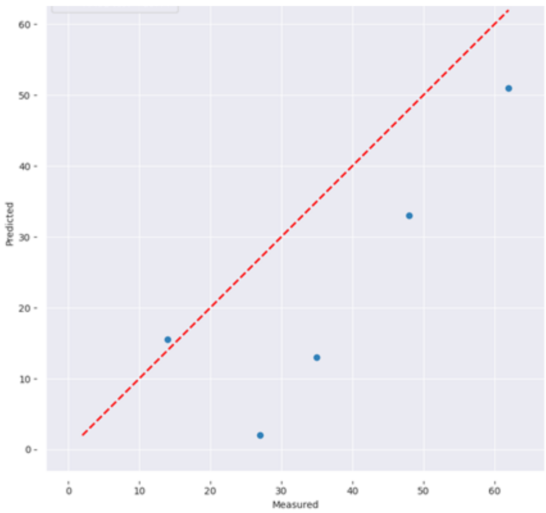

Fig. 5. The predicted RUL (in days) vs. ground truth RUL (in days) for the proposed stacking regressor model

To have a better sense of results, the numerical values of the RUL predicted by each algorithm are reported in Table. 7 The proposed model has better performance than RVR and linear Ridge. REXGB is overestimated at the end of the bearing life, which is not acceptable. So, the best results have been achieved by the proposed model.

Like all methods, the stacking process faces some issues. One of the major issues in the stacking regressors is training time which can be so large compared to a single algorithm. It causes serious issues while dealing with large datasets. Since this paper aims to suggest a method for predicting the RUL based on a limited number of points it cannot be a problem. Furthermore, the thing that should be noted is that the process of choosing base and meta-models is almost done by experience. Therefore, it demands further studies to investigate the different techniques by which the performance of stacked models is improved.

Table. 7. Comparison of the predicted RUL with actual values.

Model	**Prediction Point (Day)**				
Actual	62	48	35	27	14
RVR	25	19	9.3	10	4
REXGB	51	45.6	13	1.1	25
Linear Ridge	34	25	12	3	11.3
The proposed model (stacking regressor)	51	33	13	2	15.5

Summary/Conclusion

This paper proposed a novel method in the field of RUL prediction of ball bearings which increases the reliability of the prediction and can be trained with a limited number of vibration signals. Since most industrial plants use offline condition monitoring techniques due to financial issues, the proposed model is practical. The model is a stacking regressor with three based models, including RVR, REXGB, and linear Ridge, and a linear Ridge meta-model that combines the three based models' results. The RMS and peak values of the vibrational acceleration of the bearings in the vertical direction have been considered input features. These features are recorded in most industries for critical machines. The implementation of the model on industrial data illustrated the underestimated behavior of the model that estimates the bearing RUL about 15 days earlier than failure. The superiority of the stacked model's results over the single models has been shown. The stacking regressor has been built on the most conventional machine learning models. Future interests will be changing the base models and evaluating the results for RUL prediction of other industries' ball bearings.

Acknowledgment

The authors wish to thank Imam Khomeini oil refining company of Shazand for proving the industrial data.

References

[1] Mahamad, A.K.; Saon, S.; Hiyama, T. "Predicting remaining useful life of rotating machinery based artificial neural network". *Comput. Math. Appl.* **2010**, *60*, 1078–1087.

[2] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, "Remaining useful life estimation based on nonlinear feature reduction and support vector regression," Engineering *Applications of Artificial Intelligence*, Volume 26, Issue 7, 2013

[3] Murtagh, F. "Multilayer perceptrons for classification and regression". *Neurocomputing*, 1991, 2, 183–197.

[4] Theodoropoulos, P.; Spandonidis, C.C.; Giannopoulos, F.; Fassois, S. A Deep Learning-Based Fault Detection Model for Optimization of Shipping Operations and Enhancement of Maritime Safety. *Sensors 2021*, 21, 5658

[5] Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech. Syst. Signal Process.* 2018, 104, 799–834

[6] Tian, Z. "An artificial neural network method for remaining useful life prediction of equipment subject to

condition monitoring”. *J Intell Manuf* 23, 227–237 (2012). https://doi.org/10.1007/s10845-009-0356-9

[7] M. Behzad, **H. A. Arghand**, A. Rohani Bastami,” Remaining useful life prediction of ball-bearings based on high-frequency vibration features Proceedings of the Imech, Part C: *Journal of Mechanical Engineering Science* (2018)

[8] Mehdi Behzad, Amirhossein Mollaali, Motahareh Mirfarah, Hesam Addin Arghand, “Investigation on the effects of measurement and temporal uncertainties on rolling element bearings prognostics,” *Journal of Theoretical and Applied Vibration and Acoustics, (2020), 1-16, 6(1)*

[9] Behzad, M.; Feizhoseini, S.; Arghand, H.A.; Davoodabadi, A.; Mba, D., “Failure Threshold Determination of Rolling Element Bearings Using Vibration Fluctuation.*”, Applied Sciences, 2020*

[10] Haozhe Zhang, Dan Nettleton, Zhengyuan, “Regression-Enhanced Random Forests,”*Section on Statistical Learning and Data Science, 2017*