Industrial Duct Fan Maintenance Predictive Approach based on Random Forest

Mashael Maashi, Nujood Alwhibi, Fatima Alamr, Rehab Alzahrani,
Alanoud Alhamid and Nourah Altawallah

Department of Software Engineering, College of Computer and
Information Sciences, King Saud University, Riyadh, Kingdom of Saudi Arabia

ABSTRACT

When manufacturers equipment encounters an unexpected failure, or undergo unnecessary maintenance pre-scheduled plan, which happens for a total of millions of hours worldwide annually, this is time-consuming and costly. Predictive maintenance can help with the use of modern sensing technology and sophisticated data analytics to predict the maintenance required for machinery and devices. The demands of modern maintenance solutions have never been greater. The constant pressure to demonstrate enhanced cost-effectiveness return on investment and improve the competitiveness of the organization is always combined with the pressure of improving equipment productivity and keep machines running at the maximum output. In this paper, we propose maintenance prediction approach based on a machine learning technique namely random forest algorithm. The main focus is on the industrial duct fans as it is one of the most common equipment in most manufacturing industries. The experimental results show the accuracy, reliability of proposed Predictive Maintenance approach.

KEYWORDS

Predictive Maintenance, Maintenance, Random Forest, Duct Fan, Machine Learning & Artificial Intelligence

1. INTRODUCTION

Only 18% of manufacturers equipment fail due to its age, while 82% of failures occur randomly [1]. These unexpected failures cost the industrial manufacturers an estimated $50 billion each year [2].

Maintenance organizations across industries are at different stages of maturity with different maintenance approaches. Some may be running scheduled maintenance checks based on estimates or Original Equipment Manufacturers (OEM) recommendations, while others may utilize statistics-based programs individually tailored to each fixed asset. These Preventative Maintenance (PM) programs does not accommodate with the latest industry revolution (industry 4.0) [3], which stands for predicting future failures in assets before occurring. Predictive Maintenance (PdM) offers the potential to optimize maintenance tasks in a real time, prevent

unexpected failures, maximizing the useful life of the equipment while still avoiding disruption to operations. Also, this approach promises cost savings over preventive maintenance programs.

In recent years, several studies have been reported that used machine learning techniques to accurately predict the abnormality and remaining useful life of equipment. Chigurupati et al[4] used the Support Vector Machine (SVM) model with a Gaussian Kernel for their approach with a dataset containing 14 hardware samples. Paolanti and Romeo et al[5] Investigated the machine learning approaches for PdM. The study used Decision Forest (DF) classifier algorithm for predicting different machine states with accuracy of 95% on a dataset containing 530731 record. Mathew and Toby et al[6] constructed machine learning models based on the datasets from turbo fan engine data from the Prognostics Data Repository of NASA. The data collected from the engine has 21 sensors to collect different measurements related to the engine state at runtime. The study used ten different algorithms and compared between them. The random forest algorithm generated the least error.

In this work, the authors propose a maintenance prediction approach for industrial duct fans. The proposed method consists of three phases: the data collection phase, the data preparation phase and finally the training and prediction phase. The prediction phase consists of building a classifier to predict the abnormality and building a regression model to predict the remaining useful life.

The rest of the paper is organized as follows. Section 2 presents the proposed approach. Section 3 discusses the experimental results of the proposed approach. Finally, Section 4 provides concluding remarks.

2. THE PROPOSED APPROACH

The prediction approach proposed has a workflow as shown in Figure 1. The workflow consists of three phases: data collection phase, data preparation phase and predictive modelling phase. The extracted features from the collected data are used to build the abnormality classifier and the remaining-useful life regressor. Each of these stages is discussed with more details in the following subsections.
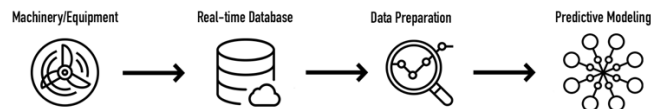

Figure 1. Predictive maintenance workflow

2.1. Data Collection

In the data collection process, MPU-6050 sensor was used to collect the data, it contains an inexpensive 3-axis accelerometer and a temperature sensor [7]. The MPU-6050 is attached to the industrial duct fan to collect the vibration and temperature data as the fan running to predict when the fan failure. Then it sends these data to the Raspberry Pi which sends the data with timestamp to a real-time database.

The output readings are sent to the real-time database every 5 seconds. The data contains the sensor ID, timestamp, x, y and z from the 3-axis accelerometer and the temperature in Celsius. The data was collected for normal and abnormal conditions. For the abnormal conditions, the authors considered some of the modes for simulating failure: increasing the temperature, deform the fan's blades shape, obstructing the movement of the fan's blades and finally simulating failure by blocking the airflow. Table 1 shows each factor with its normal and abnormal data which were taken from Duct fan model user manual [11], it's associated with failure simulation/ experiment.

The authors conduct a set of experiments in order to collect accelerometer and temperature data in failure modes. each experiment has duration of 15 to 20 minutes. Table 2 shows each experiment along with its start time, when it end and the caption of losing if it happens. Losing data happens due to voltage source is unconnected to the power. Three fan faults are shown in Figure 2 as an example.

Table 1. Duct Fan Condition

Factor		Normal Data		Abnormal Data		Failure Simulation/ Experiment
		Min	Max	Below	Above	
Temperature	Ambient	____	60 / 140	___	(+) 60° c	Increase the temperature of the room
Motion Accelerometer	External	Normal vibration data readings	Normal vibration data readings	Abnormal vibration data readings		Blocking the airflow
						Blade shape deformation
						Tape on blades

Table 2. Experimental Results of Failure Simulation

| Failure Simulation/ Experiment	Start Time	Losing Data	Return Sending	End Time	Duration	Sensor	Success
	H.M.S	H.M.S	H.M.S	H.M.S	(Min)		
Increase the temperature of the room	22.32.49	22.32.49	22.39.22	23.08.58	15	MPU 0001	yes
Blocking the airflow	22.08.10	22.10.05	22.16.36	22.29.37	15	MPU 0001	yes
Blade shape deformation	21:29:40.	21:44:20.	21:49:50.	21:56:51	20	MPU 0002	yes
Tape on blades	17:59:08	----------	----------	18:19:15	20	MPU 0002	yes
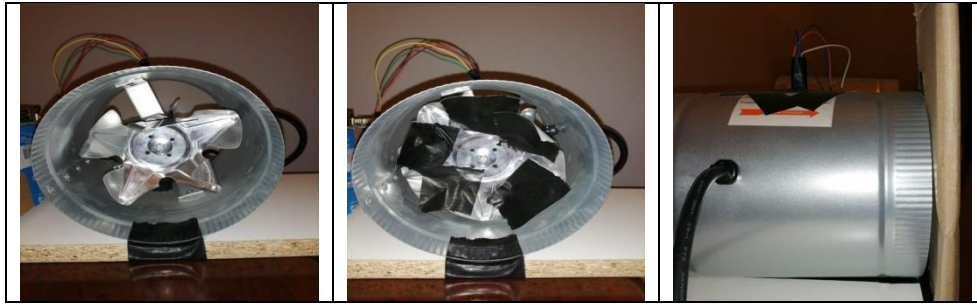
Figure 2. Faults on duct fan.

2.2. Data Preparation

The authors collect 10,963 records from the data collection process; the data needs to be transformed into a format conducive for Machine Learning. The most common is a data frame where rows represent examples to predict or learn from, and columns represent variables including features and targets. The target variables the authors considered were Abnormality and Remaining Useful Life. .The collected data were spit into training and testing set. The training set contains a known output and the model learns on this data in order to be generalized to other data. The authors have the test dataset (or subset) in order to test our model's prediction on this subset

2.3. Predictive Model

Our goal was ultimately to predict machine failures. The binary classification is viably used for predictive maintenance, being able to estimate the abnormality that the equipment has an abnormal behaviour or not. The regression models in predictive maintenance are used to calculate the remaining useful life of an asset, and it is defined as the amount of time during which the asset remains operational before the next failure occurs. The aim is to find a model that calculates the remaining useful life of each new example as a continuous number.

Random Forest algorithm (RF) [10] was applied to predict the outcomes, it is an ensemble learning method, It operates by constructing a set of decision trees at training time and outputting the mean prediction of the individual trees [8][9]. Decision trees are indeed ideal candidates for ensemble methods since they usually have low bias and high variance, making them very likely to benefit from the averaging process.

The random forests algorithm (for both classification and regression) is worked as follows:

1- Draw ntree bootstrap samples from the original data.
2- For each of the bootstrap samples, grow an un-pruned classification or regression tree, with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample mtry of the predictors and choose the best split from among those variables.
3- Predict new data by aggregating the predictions of the ntree trees (i.e., majority votes for classification, average for regression).

The structures of random forests classifiers and regressor are illustrated in Figures 3 and 4 respectively.
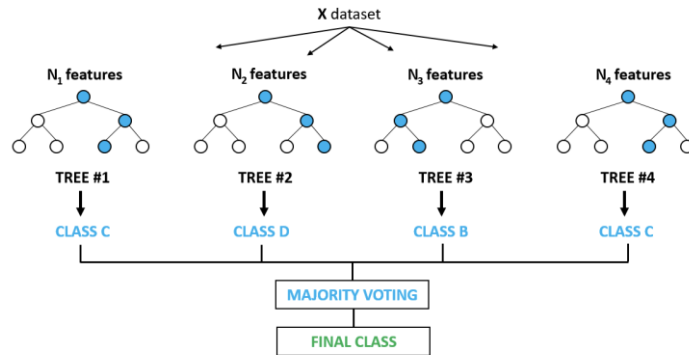

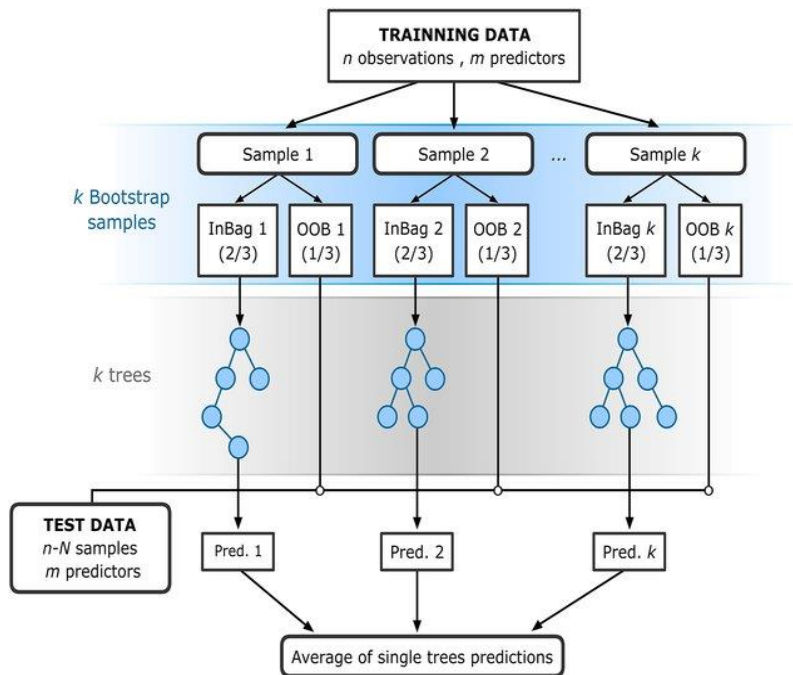Figure 3. Random forest classifier. Adopted from [12]

Figure 4. Random forest regressor. Adopted from [13]

2.4. Evaluation Measurements

The authors used the accuracy evaluation metric of the classification model by using the formula:

$$Accurcy = \frac{\text{true positives+true negatives}}{\text{total smaples}}$$
(1)

The authors used the Root Mean Squared Error(RMSE) metric in order to evaluate the

regression model , where yi is predicted value and yi^ is actual value by using the formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^{\wedge})^2} \qquad\qquad (2)$$

3. COMPUTATIONAL RESULTS

In this section, performance evaluation results of the prediction model are presented. The models were evaluated using the collected data which contains 10,963 records. The dataset was divided into training and testing dataset. The training dataset which is used in the training step includes 70% of the records and the remaining 30% is used in the testing step. According to performances metrics that used to evaluate the random forest classifier and regressor, the abnormality classifier achieved an average accuracy of 99% while the RMSE of the regressor approach achieved a value of 80.

The authors reassured that the models do not overfit by creating validation dataset and compare the model prediction against the validation labels. Figure 5 shows the model that was trained and validated on the training data. The graph produces two complexity curves, one for training and one for validation. The model does not seem to suffer from high variance and both have a high score, so the performance seems reasonably good.
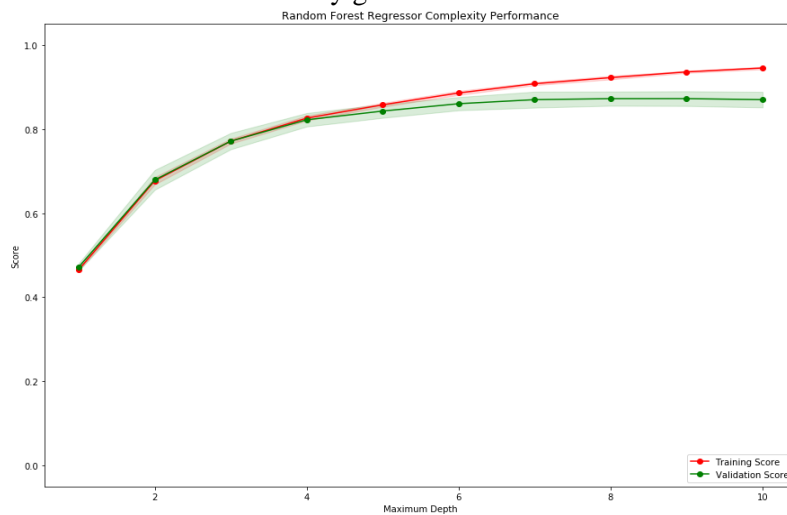

Figure 5. Complexity Performance.

4. CONCLUSION

In this paper, the authors demonstrated the potential of Machine Learning techniques on enhancing the operations of equipment and a   methodology based on PdM machine learning approach on a duct-fan machine is presented. The methodology has been implemented in a real experimental environment and data has been collected by various sensors. Future work will go in the direction of having more robust dataset, investigating different fault scenarios, exploring a different set of features and add more relevant sensors.

REFERENCES

[1] Arcweb.com. (2018). Enterprise Asset Management (EAM) Market Size, Trends, Analysis, Forecast | ARC Advisory Group. [Online] Available: https://www.arcweb.com/market-studies/enterprise-asset-management [Accessed: 15- Dec- 2019].

[2] Emerson-Unlocking Topn Performance. (2017). Paid Program: How Manufacturers Achieve Top Quartile Performance. [Online] Available: https://partners.wsj.com/emerson/unlocking-performance/how-manufacturers-can-achieve-top-quartile-performanc [Accessed: 15- Dec- 2019].

[3] M. Mulders and M. Haarman, "Predictive Maintenance 4.0 Predict the unpredictable", PwC.ln, 2017.[Online]. Available: https://www.pwc.nl/nl/assets/documents/pwc-predictive-maintenance-4-0.pdf. [Accessed: 15- Dec- 2019].

[4] A. Chigurupati, R. Thibaux, and N. Lassar."Predicting hardware failure using machine learning".Annual Reliability and Maintainability Symposium, IEEE, 2016.

[5] M. Paolanti, L. Romeo, A. Felicetti, A. Mancini, E. Frontoni, and J. Loncarski. "Machine Learning approach for Predictive Maintenance in Industry 4.0". International Conference on Mechatronic and Embedded Systems and Applications, IEEE, 2018.

[6] V. Mathew, T. Toby, V. Singh, B.M. Rao, and M.G. Kumar. "Prediction of Remaining Useful Lifetime (RUL) of turbofan engine using machine learning".International Conference on Circuits and Systems, IEEE, 2018.

[7] MPU 6050 Datasheet, [Online]. Available:https://octopart.com/mpu-6050-invensense-19505926# .[Accessed November. 2, 2019]

[8] C. Jennings, D. Wu, J. Terpenny, "Forecasting obsolescence risk and product life cycle with machine learning", IEEE TRansactions on Components Packaging and Manufacturing Technology, vol. 6, no. 9, pp. 1428-1439, 2016.

[9] W. Lin, Z. Wu, L. Lin, A. Wen, J. Li, "An ensemble random forest algorithm for insurance big data analysis", IEEE Access, vol. 5, pp. 16568-16575, 2017

[10] L. Breiman, "Random Forests", Machine Learning , vol. 45, pp. 5-325, 2011

[11] iPower-Booster-Inline-Exhaust-Grounded, [Online]. Available:https://www.amazon.com/iPower-Booster-Inline-Exhaust-Grounded/dp/B008N4QIZG .[Accessed November. 2, 2019]

[12] "Random Forest Classifier – Machine Learning | Global Software Support", Globalsoftwaresupport.com, 2019. [Online]. Available:https://www.globalsoftwaresupport.com/random-forest-classifier-bagging-machine-learning/. [Accessed: 19- Mar- 2019]

[13] V. Rodriguez-Galiano, M. Sanchez-Castillo, J. Dash, P. Atkinson and J. Ojeda-Zujar, "Modelling interannual variation in the spring and autumn land surface phenology of the European forest", Biogeosciences, vol. 13, no. 11, pp. 3305-3317, 2016

AUTHORS

Mashael S. Maashi, She received the B.Sc. degree in computer Science from King Abdulaziz University, Jeddah, Saudi Arabia, in 2005, the M.Sc. degree in Computer Science from Glamorgan University, Wales, United Kingdom in 2009, and the Ph.D. degree in Computer Science from University of Nottingham, Nottingham, United Kingdom in 2014. She is currently an assistant professor in the Department of Software Engineering, College of Computer Science and Information sciences, King Saud University, Saudi Arabia. Her main research areas are, machine learning, artificial intelligence,   evolutionary computation, optimization, intelligent computing and their applications. She is Editor of the International Journal of Computational & Neural Engineering (IJCNE). And Lead Editor for special issue theme "Computational Engineering" in IJCNE. She a board member for Global 2019 Congress on Computer Engineering. She a member in Technical Program Committee TPC of many International Conferences including: Evolutionary Computing (MIC-Evolu 2019), Computer Science and Engineering (MIC-Computing-19), Big Data and Security (ICBDS 2019) and Future Communications and Computing (MIC-FutureCom 2020).