Al-Azhar Engineering 16th International Conference

Vol. 19, No. 72, July 2024, 105 - 116

A MODEL FOR ESTIMATING RESDUIAL LIFE TIME OF MECHANICAL PARTS

Abdelrahman Shaaban*, Ali Naser Abdallah, Said M.A. Ibrahim

Mechanical Engineering Department, Faculty of Engineering, Al-Azhar University, Cairo, Egypt

***Correspondence:** Abderahman.abouzied@siemens-energy.com

Citation:

A. Shaaban, A.N. Abdallah and S.M.A. Ibrahim, "A Model for estimating resduial life time of mechanical parts", Journal of Al-Azhar University Engineering Sector, vol. 19, pp. 105 - 116, 2024.

Received: 19 November 2023

Revised: 08 January 2024

Accepted: 15 January 2024

DOI:10.21608/auej.2024.252048.1496

ABSTRACT

RUL is very important for mechanical systems, in order to predict the remaining time to failure. This is essential for safety, maintenance cost and loss of time. This would be more pronounced in turbojets in aero planes and the like. This research presents a methodology to predict the residual lifetime and technical defect of mechanical components. The present model is applied, as a case study, to data supplied by NASA of a real turbojet engine (Williams FJX-2), in order to validate the model. The dataset was divided into dependent and independent variables. Data visualization and feature selection are used to make the model more accurate by utilizing NumPy, matplotlib, Pands and Seaborn libraries via Python Programming. The correlation between variables is used to build up the model. The present model is a linear regression one. The linear regression model simplifies calculations, and most importantly, linear equations make it easy-to understand interpretation on a modular level. Though, simple but provides accurate results. To overcome the increase in the root mean square values of results and to increase the model accuracy, the linear regression is coupled with Convolutional Neural Networks (CNN); this merge improved the results greatly. The model results indicated very good agreement with real data of the turbojet engine, which gives confidence in the present model. The mathematical solution is capable to estimate the residual time for any mechanical component and not limited to the current case study.

KEYWORDS: residual lifetime, Turbofan engine, Linear Regression model, CNN algorithms.

نموذج لتقدير العمر الإنتاجي المتبقي للمكونات الميكانيكية

عبدالرحمن شعبان*، على ناصر، سعيد محمد على ابراهيم

قسم الهندسة الميكانيكية، كلية الهندسة، جامعة الأزهر، مدينة نصر 11884، القاهرة، مصر

***البريد الاليكتروني للباحث الرئيسي :** Abderahman.abouzied@siemens-energy.com

الملخص

يعد RUL مهمًا جدًا للأنظمة الميكانيكية، من أجل التنبؤ بالوقت المتبقي للفشل. وهذا أمر ضروري للسلامة وتكاليف الصيانة وضياع الوقت. سيكون هذا أكثر وضوحًا في المحركات النفاثة في الطائرات وما شابه. يقدم هذا البحث منهجية للتنبؤ بالعمر المتبقي والعيوب الفنية للمكونات الميكانيكية. تم تطبيق النموذج الحالي، كدراسة حالة، على البيانات المقدمة من وكالة ناسا لمحرك نفاث حقيقي (Williams FJX-2)، من أجل التحقق من صحة النموذج. تم تقسيم مجموعة البيانات إلى متغيرات تابعة ومستقلة. يتم استخدام تصور البيانات واختيار الميزات لجعل النموذج أكثر دقة من خلال استخدام مكتبات NumPy وmatplotlib وPands وSeaborn

عبر برمجة Python. يتم استخدام الارتباط بين المتغيرات لبناء النموذج. النموذج الحالي هو نموذج الانحدار الخطي. يعمل نموذج الانحدار الخطي على تبسيط العمليات الحسابية، والأهم من ذلك، أن المعادلات الخطية تجعل من السهل فهم التفسير على المستوى المعياري. على الرغم من أنها بسيطة ولكنها توفر نتائج دقيقة. للتغلب على الزيادة في جذر متوسط القيم المربعة للنتائج ولزيادة دقة النموذج، يقترن الانحدار الخطي بالشبكات العصبية التلافيفية (CNN)؛ أدى هذا الدمج إلى تحسين النتائج بشكل كبير. أشارت نتائج النموذج إلى توافق جيد جداً مع البيانات الحقيقية للمحرك النفاث، مما يعطي الثقة في النموذج الحالي. الحل الرياضي قادر على تقدير الوقت المتبقي لأي مكون ميكانيكي ولا يقتصر على دراسة الحالة الحالية.

الكلمات المفتاحية: العمر الافتراضى المتبقى، المحرك المروحى النفاث، خوارزميه الانحدار الخطى، الشبكة العصبية الترشيحية.

1. INTRODUCTION

Different models are used for the prediction of RUL om mechines and mechanical components. One of these and popularly used is the linear regression model. The variables here are one independent (predictor or explanatory) and the other dependent. The dependent variable changes with variations in the independent variable. The regression model predicts the value of the dependent variable, which is the response or outcome variable being analyzed or studied. It is a statistical predictive tool used in machine learning and data science. Machine Learning (ML) is one of the booming technologies across the world that enables computers/machines to turn a huge amount of data into predictions. Prognostic algorithms used in recent years [1,2]. this model-based methods and data-driven methods [3–5] used physical model by supporting recurrent network algorithms[6]. Zhu et al [7] using convolutional neural network. Ren et al [8] presented deep neural networks (DNNs). As the long short-term memory (LSTM)neural network can be used [9,10]. The previous methods can improvement results by single operating parameter. the change at operating condition is related to practice applications, Wu et al [11], Jiang et al [12], Rigamonti et al [13], and Yan et al [14] predicted the RUL by health indicators algorithms Although Sameer et al [15] and Tao et al [16] discovered new methods suitable for different operating condition by using the previous data [17–21], which is easy to calculate, was big used to calculate the RUL of the mechanical system under different conditions [22–23]. The STM is an good method to deal with the various conditions.

The present work focuses on developing a model based on linear regression merged with CNN to predict the residual lifetime of a real NASA turbojet engine. Public datasets to build up the model and to validate it. The model results almost coincide with the provided NASA data, which emphasis the validity of the present model.

2. Model Building

The present model is a supervised model is a sloped best fit straight line that relates the independent and dependent (output) variables.

The starting point in building the algorithm model is to collect and prepare the right ML dataset. The dataset should be of high quality and sufficient number. This is one of the most crucial steps in training a ML model. It can be the determinant between the success and failure of the model. After collecting the dataset, it is important to preprocess it by cleaning and completing it, as well as annotate the data by adding meaningful tags readable by a computer. The key purposes of the

dataset are: train the model, measure the accuracy of the model once it is trained, and improve the model in real application.
Python software is employed in this work. Python is a universal programming language found in a variety of different applications due to its ubiquity and ability to run on nearly every system architecture. Python is an interpreted, interactive, object-oriented language. It incorporates modules, exceptions, dynamic typing, very high level dynamic data types, and classes. Python offers multiple great graphing libraries packed with lots of different features. NumPy, Pandas, Seaborn, and Sklearn are a few of the foremost prevalent libraries utilized in Python programming. NumPy may be a library for scientific computing, Pandas could be a library for data analysis, Seaborn could be a library for visualizing information, and Sklearn could be a library for machine learning. Matplotlib is a visualization library in Python for 2D plots of arrays, It is a multi-platform data visualization library built on NumPy arrays and can work with SciPy stack. Each library provides effective, however simple, data manipulation and analysis tools. With these libraries, one could rapidly and effectively make capable applications that use the control of data science.
The dataset used in the current model is that provided by NASA for a turbofan jet engine. Figure 1 shows the components of the NASA turbojet engine under study. The engine is manufactured by WILLIAMS INTERNATIONA, USA, model Williams FJX-2. The engine's rotating components are: fan, high pressure compressor (HPC), low pressure compressor (LPC), high pressure turbine (HPT), and low pressure turbine (LPT). The gas flows through HPC, HPT, and LPT. The engine was initially tested in 1998. It is a two shaft turbofan jet engine with a combustor diameter of 14.49 in. It has a length of 40.98 in, and weights 85.98 Ib.

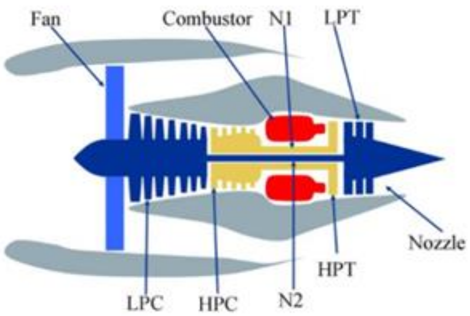


Figure 1. A simplified diagram of the turbjet.

The model is built in the following steps:
● **Step one: dataset**
Understanding the nature of data , physical cycle and technical specifications in order to make data suitable for investigation and analysis
Table 1 depicts the sensors used to monitor the operation parameters of NASA turbofan jet engine to extract the data from the system.

Tabel 1 Name of sensors.

Index	Symbol	Description	units
1	T2	Total temperature at fan inlet	°R
2	T24	Total temperature at LPC outlet	°R
3	T30	Total temperature at HPC outlet	°R
4	T50	Total temperature at LPT outlet	°R
5	P2	Pressure at fan inlet	psia
6	P15	Total pressure in bypass-duct	psia
7	P30	Total pressure at HPC outlet	psia
8	Nf	Physical fan speed	rpm
9	Nc	Physical core speed	rpm
10	epr	Engine pressure ratio (P50/P2)	–
11	Ps30	Static pressure at HPC outlet	psia
12	phi	Ratio of fuel flow to Ps30	pps/psi
13	NRf	Corrected fan speed	rpm
14	NRc	Corrected core speed	rpm
15	BPR	Bypass Ratio	–
16	farB	Burner fuel-air ratio	–
17	htBleed	Bleed Enthalpy	–
18	Nf_dmd	Demanded fan speed	rpm
19	PCNfR_dmd	Demanded corrected fan speed	rpm
20	W31	HPT coolant bleed	lbm/s
21	W32	LPT coolant bleed	lbm/s

The sensors in Table explain a function for all sensors .

● Step two: data visualization

Data visualization is the process of representing data in visual form, which could include charts, maps, and graphs. Data visualization is a powerful way to extract insights from data, as it is easier to identify outliers, patterns, and trends. Next steps 3 and 4 are related to data visualization.

After collecting the data in CSV file sheet from sensors and updating this data in the Excel sheet we need to visualize these data to calculate the average standard deviation, maximum and minimum values in order to give a quick view about data before starting the analysis to arrange the ideas in a strategy to solve the model. **T**his step used the Pandas library in python to produce Table 2.

Table 2 Data visualization as produced by Pandas library.

	id	cycle	op1	op2	op3	sensor1	sensor2	sensor3	sensor4	sensor5	...	sensor13	sensor14	sensor15
count	20631.000000	20631.000000	20631.000000	20631.000000	20631.0	20631.00	20631.000000	20631.000000	20631.000000	2.063100e+04	...	20631.000000	20631.000000	20631.000000
mean	51.506568	108.807862	-0.000009	0.000002	100.0	518.67	642.680934	1590.523119	1408.933782	1.462000e+01	...	2388.096152	8143.752722	8.442146
std	29.227633	68.880990	0.002187	0.000293	0.0	0.00	0.500053	6.131150	9.000605	1.776400e-15	...	0.071919	19.076176	0.037505
min	1.000000	1.000000	-0.008700	-0.000600	100.0	518.67	641.210000	1571.040000	1382.250000	1.462000e+01	...	2387.880000	8099.940000	8.324900
25%	26.000000	52.000000	-0.001500	-0.000200	100.0	518.67	642.325000	1586.260000	1402.360000	1.462000e+01	...	2388.040000	8133.245000	8.414900
50%	52.000000	104.000000	0.000000	0.000000	100.0	518.67	642.640000	1590.100000	1408.040000	1.462000e+01	...	2388.090000	8140.540000	8.438900
75%	77.000000	156.000000	0.001500	0.000300	100.0	518.67	643.000000	1594.380000	1414.555000	1.462000e+01	...	2388.140000	8148.310000	8.465600
max	100.000000	362.000000	0.008700	0.000600	100.0	518.67	644.530000	1616.910000	1441.490000	1.462000e+01	...	2388.560000	8293.720000	8.584800

8 rows × 27 columns

Table 2 exhibits the number of rows and columns of data which are used in this step is 8 *27. It indicates also the average standard deviation, and maximum and mini mum values for every columns. The results in the table conclude that the data is well visualized and deduced.

It is noticed that the relationships between the residual life times and all sensors are almost unchanged, so all sensors readings should be considered during the calculate procedure of the times of all components. This is illustrated further in Figs. 3 and 4 as obtained by Pandas in Python.

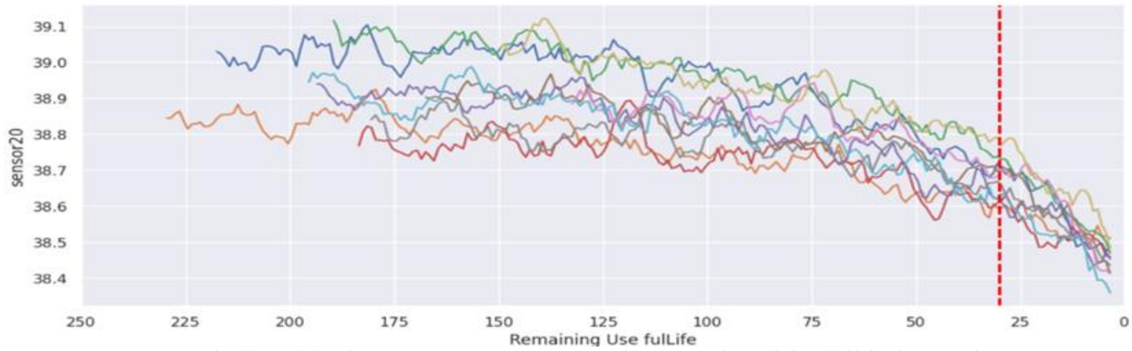


Figure 3. Relationship between sensor 20 values and residual lifetime of components.

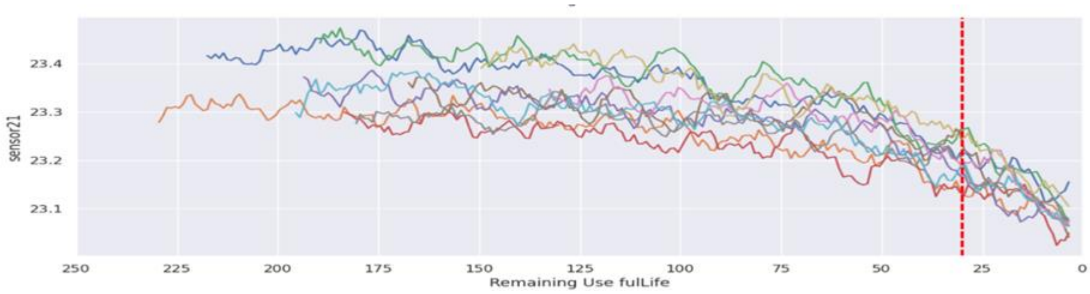


Figure 4. Relationship between sensor 21 values and residual lifetime of components.

Figures 3 and 4 show that the impacts of values of sensors 20 and 21 on the residual lifetime of components are the same. Therefore, it is concluded that the relationship between residual lifetime and all sensors are the same, so we should take into account all readings of sensors during the analysis the residual lifetime of these components.

Step 3: dependent and independent variables

In this step, the data are divided into dependent and independent variables as reported in Table 3. This is important for starting to apply the linear regression model. This enables evaluating predicted values to overcome fitting the model with data that reduce its accuracy.

Table 3 Dependent and independent variables of the model.

dependent variable	independent variables

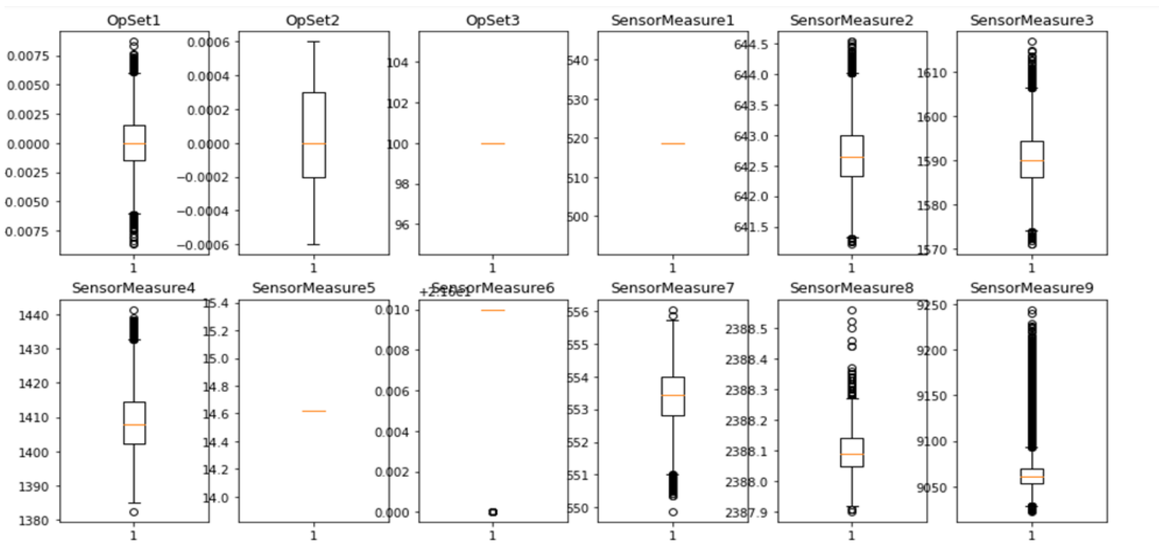
Residual lifetime	OpSet1, OpSet2, OpSet3, SensorMeasure1, SensorMeasure2, SensorMeasure3, SensorMeasure4, SensorMeasure4, SensorMeasure5, SensorMeasure6, SensorMeasure7, SensorMeasure8, SensorMeasure9, SensorMeasure10, SensorMeasure11, SensorMeasure12, SensorMeasure13, SensorMeasure14 SensorMeasure15, SensorMeasure16, SensorMeasure17, SensorMeasure18, SensorMeasure19, SensorMeasure20, SensorMeasure21.

Step 4: correlation between data

Here we find the interconnection relationship between data and feature selection to make data more accurate for the model. NumPy, matplotlib, and seaborn libraries in Python are used in data visualization to get this correlation.

This step has two benefits: (1) to extract data to be neglected as being unbeneficial for the model, and (2) to check the correlation coefficient between each data and another.

These data visualization process enables discovering data that have good relation with residual life time and data with no relation. Figure 5, obtained by using matplotlib and seaborn libraries, shows the results of this correlation process. Reviewing the results in the figure, we observe many empty boxes and some filled ones. The empty boxes indicate that these data have no good relationship with the residual life time, and vise versa for the filled boxes. The data in the empty boxes are dropped in order not to affect learning of algorithms if are not dropped, otherwise the algorithms will be either overfitting or underfitting which must be avoided for model accuracy.



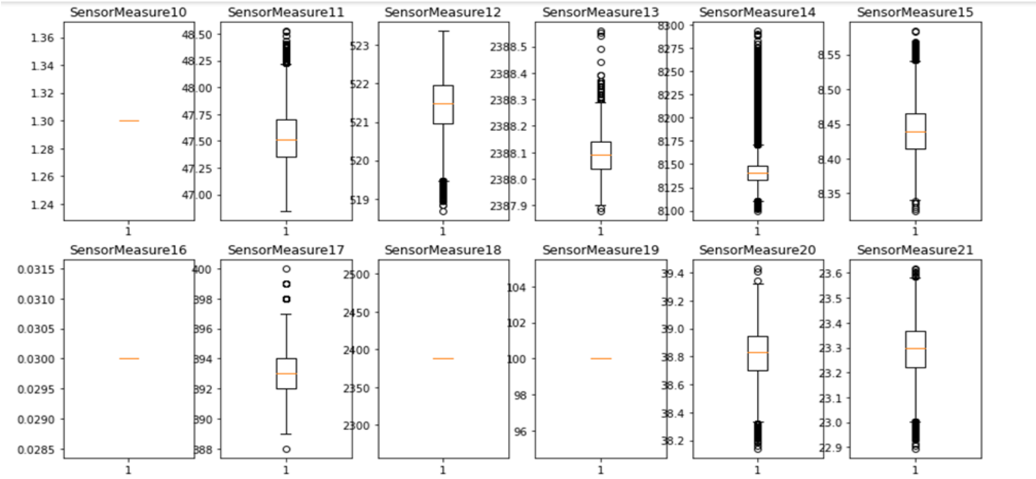


Figure 5. Useful and un-useful data for the model

Figure 5 shows that the dropped data are: 'OpSet1', 'OpSet2', 'OpSet3', 'SensorMeasure1', 'SensorMeasure5', 'SensorMeasure6', 'SensorMeasure9', 'SensorMeasure10', 'SensorMeasure14', 'SensorMeasure16', 'SensorMeasure18', and 'SensorMeasure19.

Now, the correlation between variables is checked to build up a more accurate model. The outcome of this check evaluation is indicated in Fig. 6, using NumPy and matplotlib libraries, in which the un-useful data from the previous results are dropped. The information in Fig. 6 check the correlation between data and each other to ensure that the relationship between data and the statistical coefficient range for this coefficient lies between -1 and 1 on the axis according to the variation of range taking into account all values for all colors. Figure 6 ensures this concept, where sensor 9 in red color shows a good correlation value of 1. Considering the red color of this sensor, it could be concluded that the data is more accurate and precise.

Figure 6. Correlation coefficients of data

Step 5: splitting data into train and test sets

Train test split is a model validation procedure that allows you to simulate how a model would perform on new/unseen data. To build up the model, we split the data into train and test sets and calculate the root mean square error for the model.

The data now are accurate and precise, since all useless data are dropped and from data set of the turbo fan jet engine and using the dimension reduction technique to drop any dimension of data which affects the model to avoid over and under fittings of data. The dataset is divided into three equal groups: train, test, and validation datasets. After training and testing the algorithms, root mean square error (RMSE) are determined for both train and test datasets.

3. Results and discussion

The results for the linear regression model at train and test datasets are shown in Figs. 7 and 8, respectively. In these figures the x-axis represents the number of cycles and the y-axid the residual life time. Figure 7 presents the linear regression algorithms in the train test. It is seen clearly that the average RMSE is about low of only 23, which implies a quite accurate model of high performance. This is the result of the good accurate data used in the model.

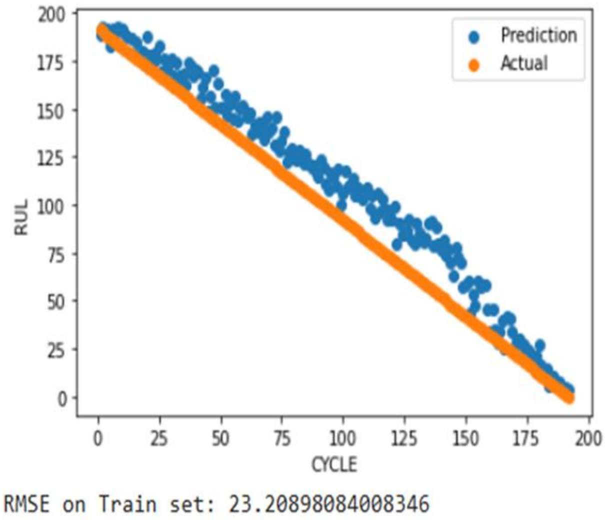


Figure 7. Linear regression algorithms at train set stage

Figure 8 shows the linear regression algorithms in the test set stage. The RMSE in the test set is a high of 48, which is more than double that of the train set. This indicates a less accurate data in this stage, which reflects adversely on the predictions at this level. Thus, the data in this stage should be improved for more accuracy and precision. It is concluded that the linear regression algorithm alone at this level is not enough to predict reliable correct values for the residual life time of the engine. The way to do this is to use using the Convolutional Neural Networks.

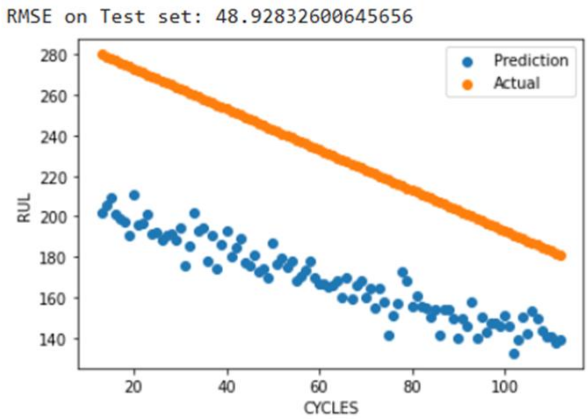
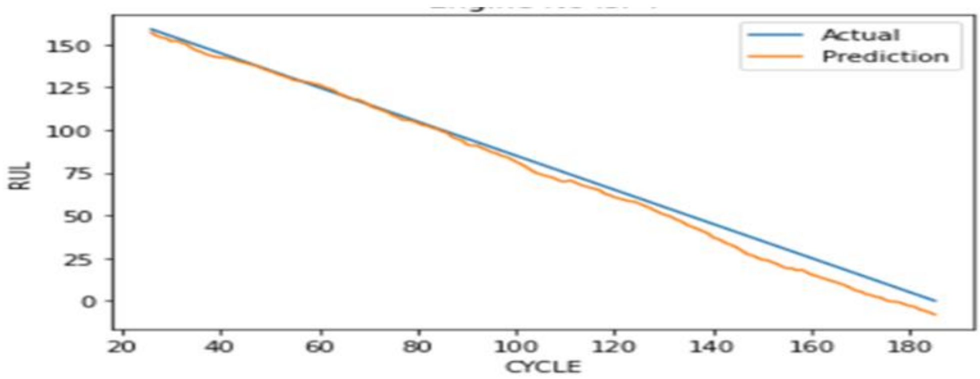


Figure 8. Linear regression algorithms at test set

Convolutional Neural Networks is employed in the present work for more model training to support the linear regression algorithm in the test set stage in order to reduce the RMSE. Figure 9 depicts the residual life time predictions of the linear regression algorithms in the test stage after applying the CNN. The RMSE is reduced tremendously to only 5. It is clear that the use of CNN improved the model accuracy a lot, since the data has passed many calculations in the internal algorithms of the CNN.

The present linear regression model coupled with CNN is accurate and reliable and proved valid for practical applications



RMSE on This set: 5.669970167381561

Figure 9. Convolutional Neural Network Algorithms at test set.

4. Conclusions

A linear regression model is developed, which makes the estimation procedure simple and, most importantly, these linear equations have an easy-to-understand interpretation on a modular level. The aims to give prediction of the life time of a NASA turbofan jet engine. The dataset is that of a NASA turbofan jet engine. These conclusions are withdrawn:

- The present linear regression model used CNN to improve its accuracy. The model proved to be accurately valid with only RMSEs of 23 and 5 in the train and test set stages, respectively.
- All readings of sensors should be considered in the analysis of failure because all sensors have the same effect on components as deduced from data visualization.
- The prediction of the life time is critical for aircraft in order to track the deterioration and failure of components during operation to detect the position of failure and give information for logistics engineer to prepare spare parts before maintenance schedule to save time and cost.
- The model incorporates high data level and analysis, in which data dimensions are reduced to overcome underfitting and overfitting of algorithms.

References

[1 Lee, J.; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L.; Siegel, D. Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. Mech. Syst. Signal Process. 2014, 4314–334.

[2]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.

[3]Li, Y.; Billington, S.; Zhang, C.; Kurfess, T.; Danyluk, S.; Liang, S. Adaptive prognostics for rolling element bearing condition. Mech. Syst. Signal Process. 1999, 13, 103–113.

[4]Li, C.J.; Lee, H. Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics. Mech. Syst. Signal Process. 2005, 19, 836–846.

[5] Oppenheimer, C.H.; Loparo, K.A. Physically based diagnosis and prognosis of cracked rotor shafts. In Proceedings of the 2nd Conference on Component and Systems Diagnostics, Prognostics, and Health Management, Orlando, FL, USA, 1–5 April 2002; pp. 122–132.

[6]Guo, L.; Li, N.; Jia, F.; Lei, Y.; Lin, J. A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing 2017, 240, 98–109.

[7]Zhu, J.; Chen, N.; Peng, W. Estimation of Bearing Remaining Useful Life based on Multiscale Convolutional Neural Network. IEEE Trans. Ind. Electron. 2018, 66, 3208–3216.

[8]Ren, L.; Sun, Y.; Cui, J.; Zhang, L. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. J. Manuf. Syst. 2018, 48, 71–77.

[9]Hinchi, A.Z.; Tkiouat, M. Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network. Procedia Comput. Sci. 2018, 127, 123–132.

[10] Zhang, Y.; Xiong, R.; He, H.; Pecht, M. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. IEEE Trans. Veh. Technol. 2018, 67, 5695–5705.

[11]Wu, Y.; Wang, Y.; Jiang, Y.; Quan, S. A Prognostic method for DC-DC converters under variable operating conditions. In Proceedings of the 2014 International Conference on Prognostics and Health Management, Cheney, WA, USA, 22–25 June 2014.

[12] Jiang, Y.; Wang, Y.; Quan, S.; Yi, W. A fault prediction method for closed-loop SEPIC converters under variable operating conditions. In Proceedings of the 2016 Prognostics and System Health Management Conference, Chengdu, China, 19–21 October 2016.

[13] Rigamonti, M.; Baraldi, P.; Zio, E.; Astigarraga, D.; Galarza, A. Particle Filter-Based Prognostics for an Electrolytic Capacitor Working in Variable Operating Conditions. IEEE Trans. Power Electron. 2015, 31,1567–1575.

[14] Yan, H.; Liu, K.; Xi, Z.; Shi, J. Multiple Sensor Data Fusion for Degradation Modeling and Prognostics under Multiple Operational Conditions. IEEE Trans. Reliab. 2016, 65, 1416–1426.

[15] Sameer, A.; Maio, F.D.; Baraldi, P.; Zio, E. Remaining useful life estimation in heterogeneous fleets working under variable operating conditions. Reliab. Eng. Syst. Saf. 2016, 156, 109–124.

[16]Tao, T.; Zio, E.; Wei, Z. A novel support vector regression method for online reliability prediction under multi-state varying operating conditions. Reliab. Eng. Syst. Saf. 2018, 177, 35–49.

[17]You, M.Y.; Meng, G. A Framework of Similarity-Based Residual Life Prediction Approaches Using Degradation Histories with Failure, Preventive Maintenance, and Suspension Events. IEEE Trans. Reliab. 2012, 62, 127–135.

[18]Zhang, Q.; Tse, W.T.; Wan, X.; Xu, G. Remaining useful life estimation for mechanical systems based on similarity of phase space trajectory. Expert Syst. Appl. 2015, 42, 2353–2360.

[19]You, M.Y.; Meng, G. Toward effective utilization of similarity based residual life prediction methods: Weight allocation, prediction robustness, and prediction uncertainty. Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng. 2013, 227, 74–84.

[20] Wang, T.; Yu, J.; Siegel, D.; Lee, J. A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems. In Proceedings of the International Conference on Prognostics and Health Management, Denver, CO, USA, 6–9 October 2008; pp. 1–6.

[21] Zio, E.; Di Maio, F. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. Reliab. Eng. Syst. Saf. 2010, 95, 49–57.

[22] Wang, P.; Youn, B.D.; Hu, C. A generic probabilistic framework for structural health prognostics and uncertainty management. Mech. Syst. Signal Process. 2012, 28, 622–637.'
[23] Li, L.L.; Ma, D.J.; Li, Z.G. Residual Useful Life Estimation by a Data-Driven Similarity-Based Approach: A Data-Driven Similarity-Based Approach. Qual. Reliab. Eng. Int. 2016, 33, 231–239.