Analysis of Failure Trends in Aircraft Engines to Predict Aircraft Engine Remaining Useful Life through Data-Driven Techniques

1. Introduction

Accurate prediction of Remaining Useful Life (RUL) in aircraft engines is of paramount importance in aviation maintenance, as it facilitates the early detection of potential failures, enabling timely maintenance or replacement of engine components. This, in turn, significantly enhances aviation safety and operational efficiency. In the context of this project, our primary focus lies on the preprocessing of sensor data and the development of a sophisticated machine-learning model for RUL prediction. The ultimate goal is to harness data-driven techniques to uncover the underlying factors that influence RUL, subsequently building predictive models to address this crucial aviation challenge. This report offers an in-depth exploration of our project's objectives and findings.

Predicting RUL is a mission-critical task in aviation maintenance, and this project leverages the extensive sensor data provided by the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset to develop and validate a model for RUL prediction. By harnessing this wealth of sensor data, our aim is to make a meaningful contribution to the enhancement of aviation safety and operational efficiency.

2. Data Preprocessing

2.1 Calculation of Remaining Useful Life (RUL)

The first step in our project involved calculating the RUL for each engine unit. This estimation was accomplished by identifying the maximum cycle count for each engine unit and subtracting the current cycle count from this maximum value.

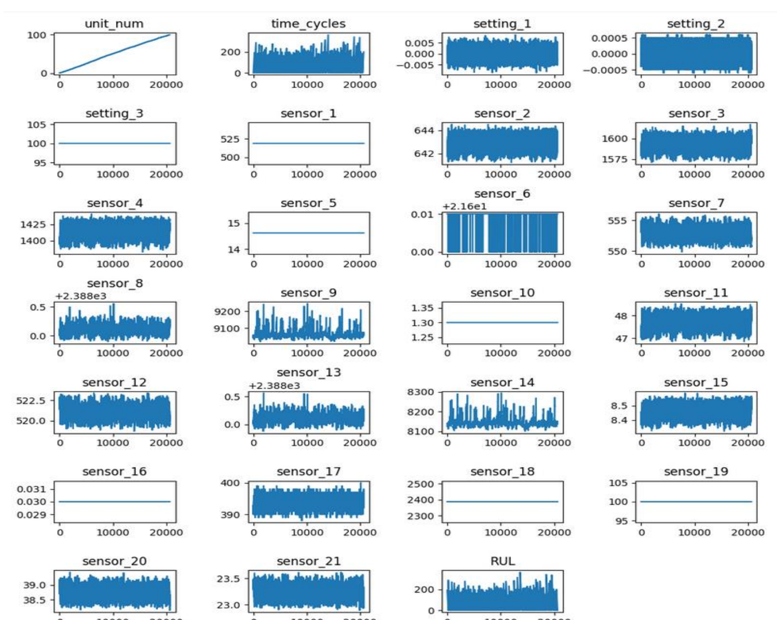

Figure 1: The x-axis represents time, while the y-axis corresponds to the variables. The significance of this visualization becomes evident as it reveals patterns, trends, and behaviors of the variables as RUL progresses.

2.2 Data Visualization

Following RUL calculation, we embarked on comprehensive data exploration. A multivariate time series plot, as shown in Figure 1, was generated to visualize the patterns exhibited by various variables in relation to RUL. Data visualization is a crucial step in enhancing our understanding of the dataset. It aids in data quality assessment, informative variable selection, and informed decision-making for subsequent modeling and analysis.

2.3 Preprocessing Steps

2.3.1 Mean and Standard Deviation Calculation:

Some sensors in the dataset exhibited constant, flat-line behavior. To identify these uninformative variables, we calculated both the mean and standard deviation for all sensors. A standard deviation close to zero indicated a lack of variability, making these sensors less valuable for RUL prediction. As presented in Table 1, Setting 3 and sensors 1, 5, 10, 16, 18, and 19 exhibited zero standard deviation, confirming their constant values. It became evident that these sensors should be excluded from our predictive modeling efforts.

Table 1: Mean and Standard Deviation of All Features

Column Name	Mean	Standard Deviation
setting_1	-8.87015e-06	0.00218731
setting_2	2.35083e-06	0.000293062
setting_3	100	0
sensor_1	518.67	0
sensor_2	642.681	0.500053
sensor_3	1590.52	6.13115
sensor_4	1408.93	9.0006
sensor_5	14.62	1.7764e-15
sensor_6	21.6098	0.00138898
sensor_7	553.368	0.885092
sensor_8	2388.1	0.0709855
sensor_9	9065.24	22.0829
sensor_10	1.3	0
sensor_11	47.5412	0.267087
sensor_12	521.413	0.737553
sensor_13	2388.1	0.0719189
sensor_14	8143.75	19.0762
sensor_15	8.44215	0.037505
sensor_16	0.03	1.38781e-17
sensor_17	393.211	1.54876
sensor_18	2388	0
sensor_19	100	0
sensor_20	38.8163	0.180746
sensor_21	23.2897	0.108251

2.3.2. Correlation Analysis:

Correlation analysis was a vital step in assessing the relationships between variables and RUL. The Spearman correlation method was chosen for its ability to capture non-linear relationships and handle ordinal data, which are common characteristics of time series data. The correlation results, visualized in a heatmap, supported our earlier observations. Sensors with a correlation coefficient of 0 with RUL, such as Setting 3, Sensor 1, Sensor 5, Sensor 10, Sensor 16, Sensor 18, and Sensor 19, were identified as unsuitable for analysis. Additionally, Setting 1, Setting 2, Sensor 6, and Sensor 14 exhibited relatively smaller correlation coefficients with RUL, indicating their lower predictive value compared to other variables. These findings guided our variable selection, emphasizing the importance of prioritizing sensors with stronger RUL correlations.

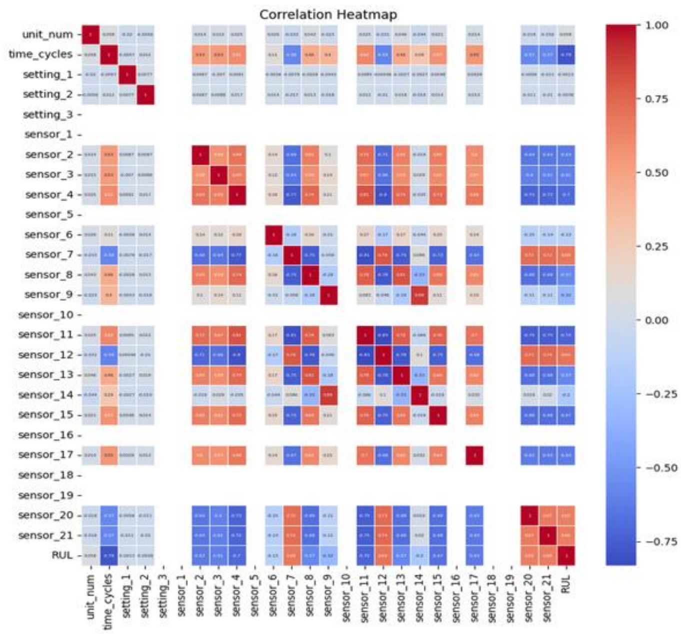

Figure 2: Correlation Heat Map

2.3.3. Granger Causality Analysis:

Granger causality analysis was employed to assess the causal relationship between two time series variables, enabling an understanding of which variables influence others. It is valuable for predicting the impact of one variable on another within a multivariate time series. The p-value in Granger causality tests signifies the statistical significance of these causal relationships. When the p-value exceeds 0.05, it indicates that the variables are not causally related, establishing a clear threshold for assessing causality. It is noteworthy that we removed sensors with constant values from this analysis, as the Granger test does not consider variables with no variation. This step is vital for precise predictive modeling.

Table 2 :. P-values from the Granger Causality Test with Lag = 1

Column Name	p-value
unit_num	0.300027
time_cycles	0.000354945
setting_1	0.677725
setting_2	0.116772
sensor_2	1.16433e-44
sensor_3	9.15533e-39
sensor_4	1.70775e-75
sensor_6	0.559062
sensor_7	2.47317e-57
sensor_8	5.84928e-47
sensor_9	2.32568e-26
sensor_11	7.26246e-72
sensor_12	1.21377e-66
sensor_13	2.89043e-51
sensor_14	1.11283e-18
sensor_15	1.40745e-51
sensor_17	3.37744e-53
sensor_20	4.49176e-60
sensor_21	2.76076e-49

2.3.4. Principal Component Analysis (PCA):

PCA was used to assess the variance explained by principal components and consider dimensionality reduction. The results revealed that the first two principal components accounted for approximately 78% of the total data variance. Importantly, there was no clear "elbow point" where explained variance significantly diminished, indicating that further dimensionality reduction would not significantly impact the dataset's richness. Therefore, we retained all variables for our analysis.

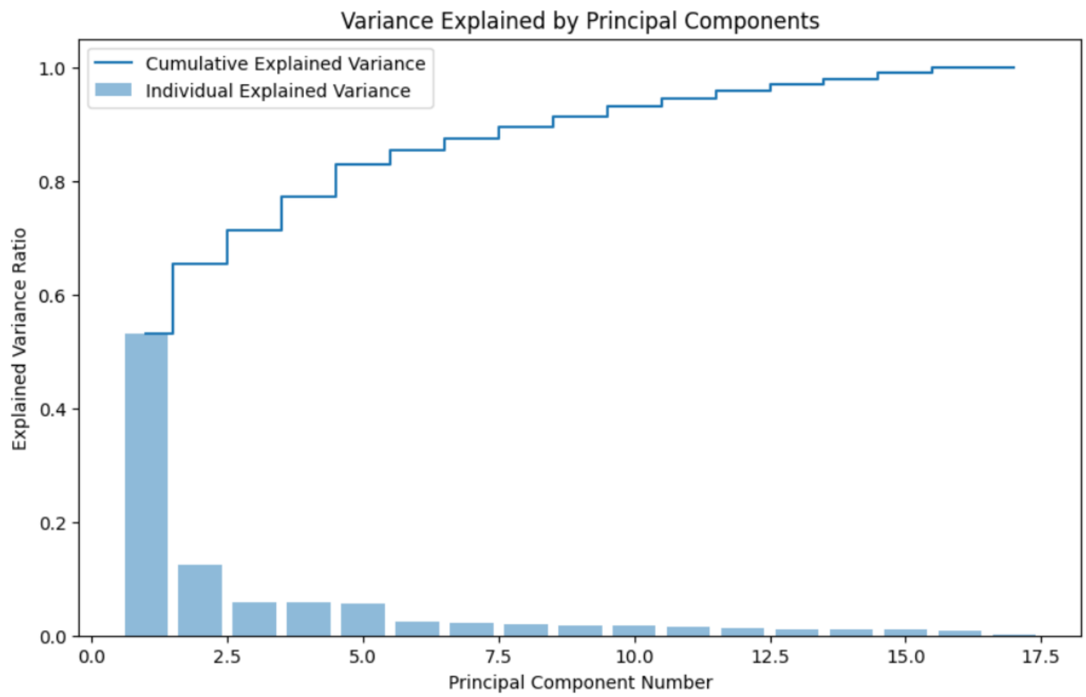

Figure 3: Variance Explained by Principal Components

Data preprocessing played a crucial role in optimizing our dataset by retaining key variables while excluding uninformative ones. This focused dataset ensures that our subsequent predictive models are built on the most relevant features. The analysis of Granger causality, Spearman correlation, and PCA enriched our understanding of the data's dynamics and relationships between variables. With this refined dataset, we are well-equipped to develop accurate RUL prediction models, enhancing the effectiveness of aircraft engine maintenance.

3. Model Development

In our modeling phase, we employed a comprehensive evaluation strategy, utilizing both Mean Absolute Error (MAE) and Mean Squared Error (MSE) as evaluation metrics. We chose this approach as each metric offers unique insights into our model's performance, addressing different aspects of our objectives. MAE provides interpretability and robustness to outliers, offering insights into the average magnitude of errors between predictions and actual values. MSE, on the other hand, is sensitive to outliers and is compatible with mathematical optimization, providing a nuanced view of our model's accuracy. By utilizing both MAE and MSE, we aimed to conduct a thorough evaluation that balances outlier-resistant performance assessment and optimization suitability, leading to a well-rounded and informed assessment of our model's capabilities. This approach accommodates the diverse characteristics and priorities in our analysis, resulting in a more robust model evaluation.

3.1 Linear Regression

Our initial modeling approach employed Linear Regression, a foundational machine learning technique that establishes a linear relationship between RUL and other sensor readings. Linear regression seeks to find the best-fitting linear equation by minimizing the sum of squared differences between predicted and observed values. The Mean Absolute Error Loss observed was 27.179516 RUL's and Mean Squared Error Loss was 27.179516 RUL's as well.

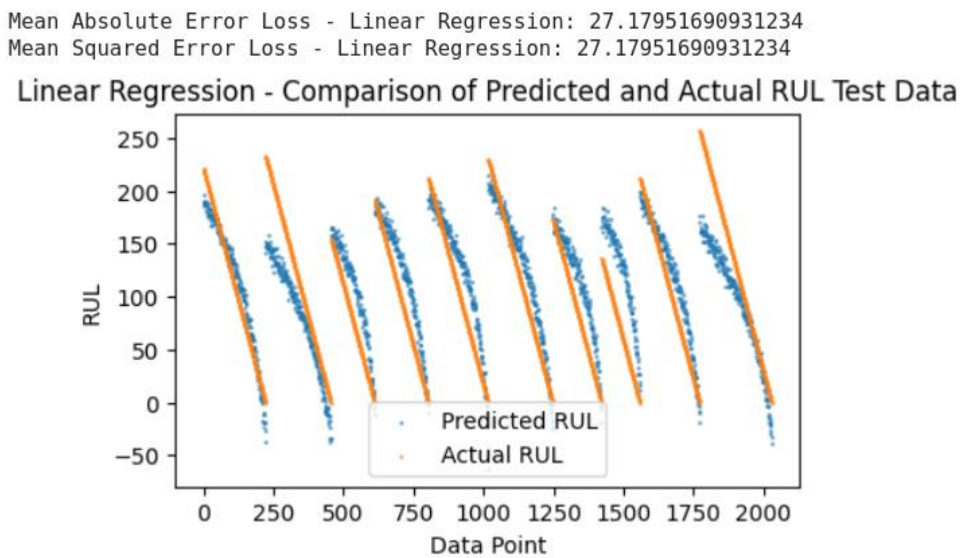

Figure 4: Linear Regression Result

3.2 Support Vector Regression (SVR)

Support Vector Regression (SVR) proved to be a versatile choice for our multivariate time series analysis, given its flexibility in capturing complex data relationships. We explored SVR with various kernel functions, including polynomial, radial basis function (RBF), and linear kernels. Notably, the linear kernel demonstrated the best performance, possibly due to the linear regression pattern present in the RUL we aimed to predict. We experimented with different regularization parameter (C) values, settling on C = 1 as it struck a balance between fitting the training data closely and model generalization. Additionally, the impact of data normalization on SVR performance was examined, revealing that normalized data consistently outperformed non-normalized data. This underscores the crucial role of data preprocessing in enhancing SVR's capabilities for multivariate time series regression, resulting in improved predictive accuracy and model generalization. The Mean Absolute Error Loss observed was 26.1129 RUL's and Mean Squared Error Loss was 1072.585 RUL.

```
Mean Absolute Error Loss - Support Vector Regression: 26.112947682034115
Mean Squared Error Loss - Support Vector Regression: 1072.585819742108
```

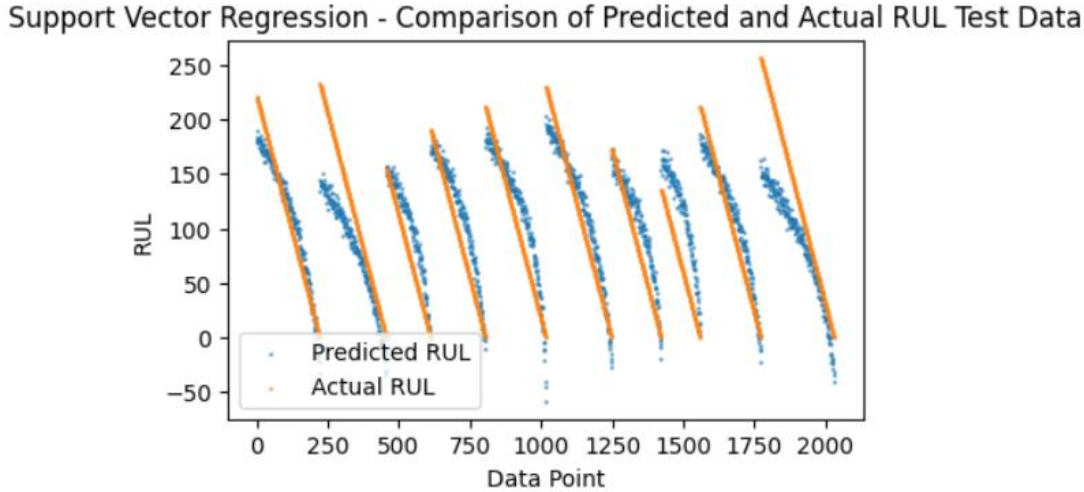

Figure 5: Support Vector Regression Result

3.3 Long Short-Term Memory) (LSTM)

Recognizing the importance of capturing the time series nature of our data, we opted for a Long Short-Term Memory (LSTM) network. Unlike traditional recurrent neural networks (RNNs), LSTMs are well-suited for sequences of varying lengths due to their ability to capture long-range dependencies and avoid vanishing gradient issues. For our LSTM architecture, we incorporated two LSTM layers with hidden sizes of 64 and 32, with batch normalization in between. This was followed by a fully connected layer. We employed the Adam optimizer with a learning rate of 0.001 for training. The decision to use LSTMs was rooted in the time series nature of the data and the need to capture temporal dependencies.

The choice of two LSTM layers in our architecture enhances the model's capacity to capture complex temporal patterns and dependencies, catering to both short-term and long-term relationships within the data. Batch normalization contributes to training stability by mitigating gradient-related issues and reducing overfitting. The Adam optimizer, with its adaptive learning rates and a rate of 0.001, combines momentum and RMSprop benefits, promoting efficient convergence and robust optimization. This approach balances model stability, speed, and accuracy for effective time series modeling.

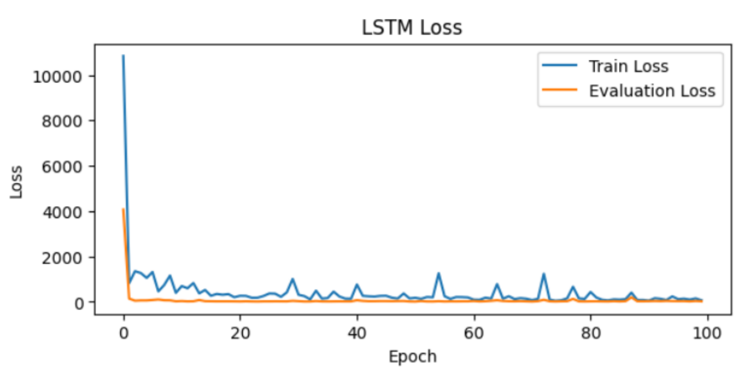

Figure 6: Training Loss and Evaluation Loss along 100 Epochs

After experimenting with sequence sizes ranging from 50 to 5, a sequence size of 30 was determined to be optimal for balancing temporal information and model complexity. Training was conducted for 100 epochs, during which the model's training loss significantly decreased, indicating effective learning. Continuous monitoring of the evaluation loss revealed that it remained stable, signifying a well-balanced approach that avoids overfitting. Our approach highlights the importance of flexible stopping criteria. By experimenting with different criteria and vigilant performance monitoring, we ensure the model is trained to capture valuable insights without unnecessary complexity. The Mean Absolute Error Loss observed was 9.4772 RUL's.

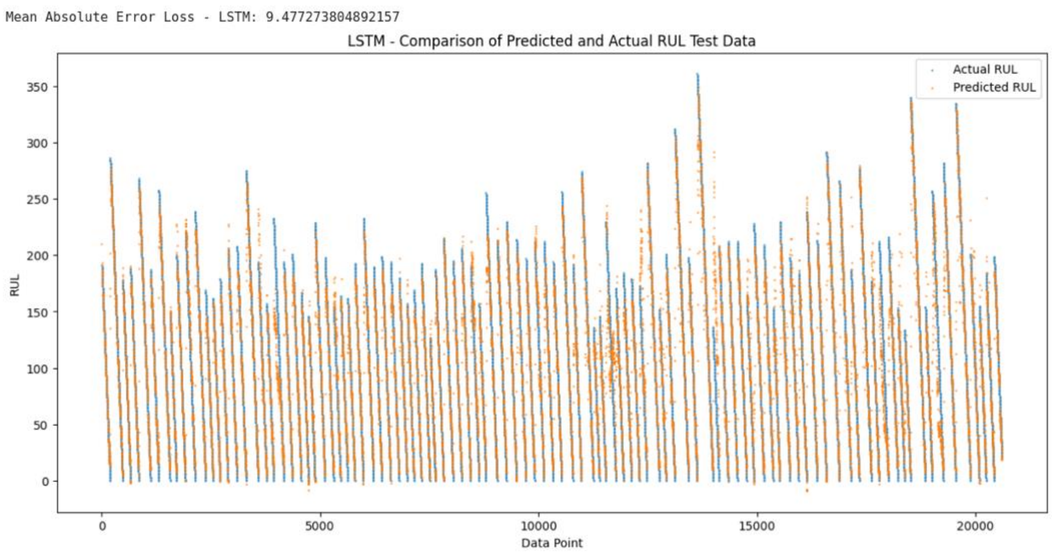

Figure 7: LSTM Result

3.4 Model Selection Rationale

The choice of models reflects a progressive refinement aimed at achieving optimal predictive accuracy. Each model offers unique strengths, addressing different aspects of the multivariate dataset and the temporal characteristics of the data.

4. Results

4.1 Post Processing

In our project, we evaluated the performance of three different algorithms for predicting Remaining Useful Life (RUL) in aircraft engines. The Linear Regression (LR) model achieved a Mean Absolute Error (MAE) of 271.179, indicating its ability to capture some aspects of the data's semi-linear shape. Similarly, the Support Vector Regression (SVR) model exhibited an MAE of 26.112 and demonstrated its capability to handle the data's linearity. However, both the LR and SVR models faced limitations when predicting RUL values, as they occasionally produced negative results, which are physically implausible in this context.

In contrast, our Long Short-Term Memory (LSTM) model outperformed the other algorithms, yielding the lowest MAE of 9.477. The LSTM model's success can be attributed to its inherent ability to capture temporal dependencies within the data. Furthermore, its robustness is highlighted by the absence of negative RUL predictions, emphasizing its superior alignment with the data's temporal dynamics.

5. Conclusion

The LSTM model's impressive performance underscores its effectiveness in modeling the complex temporal patterns of aircraft engine data. For the prediction of RUL, an essential parameter in aircraft maintenance, the LSTM model offers a superior solution. Its robustness, adaptability to time series data, and accurate predictions make it a valuable tool for the aviation industry.

6. Future Work

6.1 eXplainable Artificial Intelligence (XAI) and Post-Processing

To enhance model interpretability and transparency, we recommend the exploration of eXplainable Artificial Intelligence (XAI) techniques. XAI tools can provide insights into the LSTM model's decision-making processes, allowing for a deeper understanding of its predictions. Techniques such as SHAP (SHapley Additive exPlanations) values, LIME (Local Interpretable Model-agnostic Explanations), and feature importance analysis can help shed light on the factors driving the RUL predictions.

Additionally, post-processing steps can be applied to further refine the predictions and handle edge cases. This might involve setting a threshold to ensure that negative RUL values are automatically adjusted to zero, reflecting the practical constraints of the problem.

In conclusion, our project demonstrates the potential for advanced machine learning models like LSTM to significantly improve the prediction of aircraft engine RUL, ensuring safer and more efficient maintenance practices. By incorporating XAI techniques and thoughtful post-processing steps, we can further enhance the reliability and transparency of such models in real-world applications. This research has far-reaching implications for the aviation industry, contributing to enhanced safety, reduced maintenance costs, and improved operational efficiency.

References

Hu, Y.; Miao, X.; Si, Y.; Pan, E.; Zio, E. Prognostics and health management: A review from the perspectives of design, development and decision. Reliab. Eng. Syst. Saf. 2022, 10, 108063

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, 42, 314–334.

Jian Ma, Hua Su, Wan-lin Zhao, Bin Liu, "Predicting the Remaining Useful Life of an Aircraft Engine Using a Stacked Sparse Autoencoder with Multilayer Self-Learning", Complexity, vol. 2018, Article ID 3813029, 13 pages, 2018. https://doi.org/10.1155/2018/3813029

Wang, H.; Li, D.; Li, D.; Liu, C.; Yang, X.; Zhu, G. Remaining Useful Life Prediction of Aircraft Turbofan Engine Based on Random Forest Feature Selection and Multi-Layer Perceptron. Appl. Sci. 2023, 13, 7186. https://doi.org/10.3390/app13127186

Hai-Kun Wang, Yi Cheng, Ke Song, "Remaining Useful Life Estimation of Aircraft Engines Using a Joint Deep Learning Model Based on TCNN and Transformer", Computational Intelligence and Neuroscience, vol. 2021, Article ID 5185938, 14 pages, 2021. https://doi.org/10.1155/2021/5185938

Li H, Wang Z, Li Z. An enhanced CNN-LSTM remaining useful life prediction model for aircraft engine with attention mechanism. PeerJ Comput Sci. 2022 Aug 30;8:e1084. doi: 10.7717/peerj-cs.1084. PMID: 36091994; PMCID: PMC9455287.

Liu, L., Wang, L. & Yu, Z. Remaining Useful Life Estimation of Aircraft Engines Based on Deep Convolution Neural Network and LightGBM Combination Model. Int J Comput Intell Syst 14, 165 (2021). https://doi.org/10.1007/s44196-021-00020-1