

عبدالعزیز مرعي

thisis.docx

 Midocean University

Document Details

Submission ID

trn:oid:::2945:405103756

Submission Date

Aug 6, 2026, 8:25 PM GMT+3

Download Date

Aug 9, 2026, 4:17 AM GMT+3

File Name

thisis.docx

File Size

291.7 KB

38 Pages

6,476 Words

39,340 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely human generated as AI generated and likely AI generated as human generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.



Title

Prepared By

Name: Abdulaziz Mari

ID No.:.....

Name: Ali Alzunaydi

ID No.:.....

Submitted in fulfillment of the requirements for the degree Master of Science in Artificial
Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr

College of Informatics, Midocean University

Email:.....

Acknowledgements

Special thanks to Dr. .

Abstract

List of Publications

Table of Contents

List of Abbreviations	viii
List of Figures	Error! Bookmark not defined.
List of Tables	viii

Abstract

List of Publications

Table of Contents

List of Abbreviations xiv

List of Figures xvi

List of Tables xviii

Chapter 1: Introduction and

Overview.....	1
1.2 Problem Definition	2
1.3 Thesis Objectives	3
1.4 Thesis Contribution	4
1.5 Thesis Organization.....	5

Chapter 2: Background and Related Work
..... 31**Chapter 3: Proposed Approach** 78

4.1 Introduction

Chapter 4: Experimental Results and Discussion 86**Chapter 5: Conclusion and Future Work** 127

7.1 Important Features	127
7.3 Future Work	129
References	130

List of Abbreviations

- **Acc.:** Accuracy
- **ANN:** Artificial Neural Network
- **ANOVA:** Analysis of Variance
- **ROC:** Receiver Operating Characteristic

List of Figures

2.1 Educational Data Mining Process	7
2.2 Applications of Educational Data Mining	8
2.3 Distribution of Papers that used different attributes for PSP problem	19

2.4 Student Performance Data Set	21
2.5 Educational Process Mining (EPM): A Learning Analytics Data Set ...	23

List of Tables

- 2.1 Definitions of Traditional Machine Learning Models
- 2.2 Definitions of Deep Learning Models
- 2.3 Definitions and Equations of Evaluation Methods
- 2.4 Comparison of Related Studies

Chapter 1: Introduction and Overview

Chapter 1: Introduction and Overview

1.1 Background and Overview

Industrial equipment is used a lot in manufacturing, transportation, energy and other industries. The parts of these machines can break down over time because they are used for a time and they are affected by temperature how they are used and the environment they are in. If we do not keep an eye on this breakdown it can cause failures, high costs to fix things, safety problems and downtime. We can do maintenance in two ways: we can fix things after they break. We can do maintenance to prevent things from breaking. Preventive maintenance is done on a schedule. It can lead to unnecessary checks or replacing parts that still work. Prognostics and Health Management or PHM is a way to use sensor data to figure out the condition of equipment and decide if maintenance is needed. PHM looks at the condition of equipment checks its health and tries to predict when it will fail. One of the things PHM does is predict the Remaining Useful Life or RUL which is how much longer the equipment will work before it fails.

Machine Learning or ML and Deep Learning or DL models can look at sensor signals to learn how equipment breaks down over time. We have used models like Random Forest, Support Vector Machine. Boosting algorithms to predict when equipment will fail. DL models, like Convolutional Neural Network, Long Short Term Memory, Gated Recurrent Unit and Transformer can learn things and changes in data. We have used datasets like NASA C-MAPSS, N-CMAPSS, PHM2012 and XJTU-SY to develop models that predict when equipment will fail and to compare solutions. This thesis is about using ML and DL models to predict when equipment will fail so we can plan maintenance. Reduce the risk of unexpected failures, which makes equipment more reliable.

1.1 Background and Overview

Equipment failures can cause problems like stopping operations increasing costs to fix things, downtime and safety hazards. We can do maintenance after things break or we can do it on a schedule.. If we use sensor data to predict when equipment will fail we can get a better idea of what is going on. However predicting when equipment will fail is hard because the way equipment breaks down can be different and sensor data can be noisy.

Also things like how equipment's used, speed, environment and temperature can make it harder to make accurate predictions.

Many people have used ML and DL to predict when equipment will fail. While they got results how well the models work can depend on the data how it is prepared and the model itself. Also many people just focus on making predictions but they do not show how to use those predictions to plan maintenance. So we need a way to predict when equipment will fail that uses sensor data looks at models makes accurate predictions and gives clear instructions for maintenance. This thesis tries to do that by looking at ML and DL models and using them to predict when equipment will fail then turning those predictions into maintenance instructions.

1.2 Problem Definition

This thesis is about making a framework for maintenance that uses sensor data and ML models to predict when industrial equipment will fail. The goals are to get data from sensors look at how equipment breaks down train models evaluate them and compare them. We also want to find the model turn predictions into maintenance instructions and help maintenance managers plan what to do.

1.3 Thesis Objectives

This thesis contributes to maintenance by linking predictions of when equipment will fail to actual maintenance planning. It does this by showing the process from getting data to planning maintenance and by comparing traditional ML models to DL models. It also evaluates models turns predictions into maintenance instructions and helps maintenance managers plan ahead which reduces failures, unnecessary maintenance and downtime.

1.4 Thesis Contributions

This thesis has five chapters. Chapter 1 introduces the topic states the problem and outlines the thesis. Chapter 2 talks, about the background of maintenance PHM and RUL prediction. Chapter 3 discusses the method, including data processing.

Chapter 2: Background and Related Work

2.1 Background

PHM

Industrial machinery often shows signs of degradation before actual failure. Failure to identify the signs of machinery degradation in time can result in unexpected machinery outages, high operating and maintenance costs, and safety issues [1].

PHM is a predictive maintenance approach that relies on live sensor measurements and ML algorithms to monitor machinery health and predict failures before they occur. The primary goal of PHM is to improve equipment performance and reliability while reducing unnecessary maintenance [1].

RUL Prediction

One of the most critical tasks in PHM is RUL prediction. RUL is an estimate of the cycles, working hours or the amount of time left before a specific machinery part gets to its failure state. As the output of this task is a value which falls on a range of values, RUL prediction is categorized as a regression task. An accurate RUL prediction allows maintenance engineers to plan work equipment repairs or replacements at the ideal time, thus reducing maintenance costs and unplanned equipment outages [1].

Datasets

Several benchmark datasets are commonly used to compare RUL prediction models. The NASA CMAPSS and NCMAPSS datasets contain sensor measurements and operating condition data that were gathered from turbofan engines during their degradation process. Furthermore, the PHM2012 and XJTUSY datasets consist of vibration and temperature signals that were collected from rolling bearings under various operating conditions until a failure occurred. These datasets are widely used to benchmark RUL prediction models [1]–[15].

Data Pre-processing

Sensor data require pre-processing prior to ML training. The most frequently used pre-processing methods include noise filtering, data normalization, removal of constant sensors, RUL label generation, sliding-window segmentation, and feature extraction techniques such as Principal Component Analysis (PCA), wavelet transformation, Empirical Mode Decomposition (EMD), autoencoders, and a Health Indicator (HI) [4]–[15].

2.2 Traditional Machine Learning

Traditional RUL prediction methods include Linear Regression, SVM, KNN, Random Forest, LightGBM, CatBoost, and XGBoost. These algorithms learn the relationship between sensor measurements and equipment degradation [9].

Table 2.1. Definitions of Traditional Machine Learning Models

Model	Definition
Linear Regression	Linear Regression is a supervised model that predicts a continuous value by learning a straight-line relationship between the input features and the target value [16].
Support Vector Machine (SVM)	For regression tasks, SVM estimates continuous values by finding a function that keeps most prediction errors within a selected margin. Kernel functions can also help it learn nonlinear relationships [17].
K-Nearest Neighbors (KNN)	KNN predicts a continuous value by finding the training samples that are most similar to a new sample and combining their target values [18].
Random Forest (RF)	Random Forest trains many decision trees using different samples and features. Their outputs are then averaged to produce a more stable regression prediction [19].
Gradient Boosting	Gradient Boosting creates decision trees in sequence. Each new tree learns from the errors left by the previous trees so that the prediction improves gradually [20].
Light Gradient Boosting Machine (LightGBM)	LightGBM is a tree-based gradient boosting framework designed to train efficiently while using less memory. It can also support large datasets and parallel processing [21].
Categorical Boosting (CatBoost)	CatBoost is a gradient boosting model built with decision trees. It includes methods for handling categorical features and reducing prediction errors caused during the training process [22].
Extreme Gradient Boosting (XGBoost)	XGBoost is an optimized gradient boosting framework that combines multiple decision trees. It is designed to provide efficient training and can work with parallel and distributed computing [23].

2.3 Deep Learning

Deep learning techniques include CNN, LSTM, GRU, Bi-LSTM, CNN-LSTM, CNN-GRU, Transformer, and Attention mechanisms. Hybrid architectures generally provide better prediction accuracy because they capture both spatial and temporal degradation patterns [2], [5], [6], [8], [10]–[15].

Table 2.2. Definitions of Deep Learning Models

Model	Definition
CNN	CNN is a deep learning model that moves learned filters across the input to find useful local patterns. In predictive maintenance, it can detect degradation features from sensor and vibration signals [24].
LSTM	LSTM is a recurrent model that uses memory cells and gates to decide what information should be kept, updated, or removed. This helps it learn changes that occur across long time sequences [25].
GRU	GRU is a recurrent model that controls information through update and reset gates. Its structure is simpler than LSTM, but it can still learn how sensor patterns change over time [26].
Bi-LSTM	Bi-LSTM uses two LSTM networks. One reads the sequence forward, while the other reads it backward. Their outputs are combined to provide more sequence information [27].
CNN-LSTM	CNN-LSTM uses CNN to find local features and LSTM to study how those features change with time [24], [25].
CNN-GRU	CNN-GRU combines CNN feature extraction with GRU sequence learning to follow signal patterns across time [24], [26].
Autoencoder	An Autoencoder learns to reproduce its input through a smaller internal representation. This can reduce data dimensions, remove noise, and extract useful features [28].
Attention	An Attention Mechanism gives different importance to different input parts so the model can focus on useful features or time steps [29].
Transformer	A Transformer is a sequence model based mainly on self-attention. It examines relationships between different positions without depending mainly on recurrent processing [30].
TCN	TCN is a convolution-based model for sequential data. It often uses causal and dilated convolutions to learn short and long temporal patterns [31].

2.4 Evaluation Methods

Table 2.3. Definitions and Equations of Evaluation Methods

Evaluation Method	Definition	Equation	Interpretation

RMSE	RMSE calculates the typical difference between the actual RUL and the predicted value. Large prediction errors have a stronger effect [32].	$RMSE = \sqrt{[(1/n) \sum_{i=1}^n (y_i - \hat{y}_i)^2]}$	A smaller RMSE means closer predictions.
MAE	MAE calculates the average prediction error size without considering the error direction [33].	$MAE = (1/n) \sum_{i=1}^n y_i - \hat{y}_i $	A smaller MAE indicates lower average error.
R ²	R ² shows how much variation in actual RUL values is explained by the model [34].	$R^2 = 1 - [\sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2]$	A value closer to 1 indicates better fit.
RUL Score	RUL Score uses an asymmetric penalty where late predictions receive a larger penalty [35].	$Score = \sum_{i=1}^n s_i$	A smaller score indicates better performance.

Model performance is commonly evaluated using RMSE, MAE, R², and RUL Score. Lower RMSE and MAE indicate better prediction accuracy, while higher R² indicates better model fitting [1]–[15].

2.5 Workflow Diagram

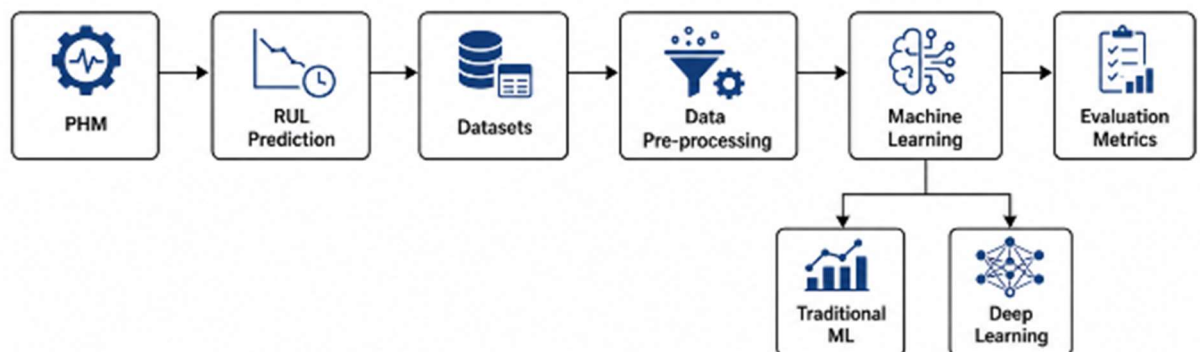


Figure 2.1. PHM-Based RUL Prediction Workflow

2.6 Related Work

Deng and Zhou [1] presented a hybrid CNN–LSTM model based on a deep residual network, LSTM and an attention mechanism for predicting the RUL of aero-engines. In their proposed model, a deep residual network was adopted as the feature extractor, an LSTM was used to observe the degradation evolution over time, and an attention mechanism was added to make the model focus on the more important areas of the sequence. NASA C-MAPSS dataset was used, which includes four sub-datasets, FD001, FD002, FD003, and FD004. The dataset contains the sensor measurements collected from aircraft turbofan engines in the course of operation until degradation and failure. In addition, the dataset includes 24 columns: the engine number, cycle number, 3 operational settings, and 21 sensor readings. Different pre-processing steps were used: normalization of sensor values, generation of RUL labels and transformation of sensor data into time-series. The model was compared with CNN, LSTM, Back Propagation neural network (BP), GRU, CNN–GRU, Deep Convolutional Neural Network (DCNN), and TCN. Results showed that CNN-LSTM performs best with RMSE of 15.977 on FD001, 14.452 on FD002, 13.907 on FD003 and 16.637 on FD004.

Fan et al. [2] proposed a Sensor-Time Attention-based RUL prediction model (STAR) based on a two-stage attention hierarchical Transformer for predicting the RUL of turbofan engines. The study used the NASA C-MAPSS dataset STAR used two attention stages to learn both sensor-wise information and temporal information It also used a hierarchical encoder-decoder structure to capture useful information at different time scales The proposed model was compared with several existing RUL prediction

models, including CNN-based models, Recurrent Neural Network-based models (RNN-based models) and Transformer-based models such as Dual-Aspect Self-Attention Transformer (DAST), DLformer and BiLSTM-DAE-Transformer. The results showed that STAR achieved an RMSE of 10.61 on FD001 13.47 on FD002 10.71 on FD003 and 15.87 on FD004 It also achieved a Score of 169 on FD001 784 on FD002 202 on FD003 and 1449 on FD004 STAR achieved the best overall results compared with the other models.

Elsherif et al. [3] proposed a hybrid deep learning model called Convolutional Autoencoder and Attention-Based LSTM (CAELSTM) to predict the RUL of turbofan engines. The study used the NASA C-MAPSS dataset, focusing on FD001 and FD003. The preprocessing steps included preparing the two sub-datasets and applying a piecewise linear degradation model. The model used a convolutional autoencoder to extract useful features from the sensor data. An LSTM layer with an attention mechanism was then used to learn the degradation pattern over time and focus on the most important parts of the sequence. A fully connected layer was used before the final RUL prediction. The model was compared with other RUL prediction approaches. Results showed that CAELSTM achieved RMSE values of 14.44 on FD001 and 13.40 on FD003. It also achieved MAE values of 10.49 on FD001 and 10.68 on FD003, with Score values of 282.38 and 264.47.

Ensarioğlu et al [4] proposed a deep learning procedure for the RUL estimation of turbofan engines using change point detection based labeling and feature engineering. The study employed the NASA C MAPSS FD001 dataset which contains simulated sensor measurements from turbofan engines operating until failure. The preprocessing procedure involved selecting 14 useful sensors removing constant sensors applying min max normalization filtering noisy signals using a low pass filter extracting a single signal using an autoencoder and constructing a new difference based feature from sensor measurements. The change point detection method was used to generate engine specific piecewise linear RUL labels instead of using the same fixed initial RUL value for all engines. Within the proposed architecture the 1D CNN component was used to extract spatial features from the prepared sensor data and the LSTM component was used to capture temporal degradation information. The results showed that the 1D CNN LSTM

model with the difference feature achieved the best prediction performance with RMSE 16.1 and Score 437.2 compared with traditional machine learning methods and earlier deep learning methods.

Sun et al. [5] presented an end-to-end model called Multi Scale Depth Wise Separable Convolution Transformer (MDSCT) for predicting the RUL of rolling bearings. The model uses (MDSC) to extract local degradation features from vibration signals and an enhanced Transformer encoder called PPSformer to learn long term relationships. An adaptive activation function called Adapth Swish was also used to improve the learning of nonlinear degradation patterns. The study used the PHM2012 and XJTU SY bearing datasets. Both contain vibration signals collected from rolling bearings during operation until degradation and failure. The XJTU-SY dataset includes 15 bearings tested under different operating conditions with horizontal and vertical vibration signals. The model used raw vibration signals as input instead of relying on many manually designed features. MDSCT was compared with Soft-Threshold Attention Mechanism-based Temporal Convolutional Network (TCN-SA), Deep Adversarial Network (DAN), Temporal Convolutional Network with Separable Residual Blocks (TCN-RSCB), Conditional Variational Transformer (CVT), Convolutional Long Short-Term Memory Network (CLSTM), Multi-Scale Convolutional Neural Network (MCNN), Self-Attention Mechanism-based Adaptive Graph Convolutional Network (SAGCN-SA), and Deep Convolutional Neural Network (DCNN). Results showed that MDSCT achieved the best average performance on PHM2012 with an MAE of 0.100 and RMSE of 0.124. On XJTU-SY, it achieved an average MAE of 0.134, an average RMSE of 0.160, and an average R^2 value of 0.667.

Jin et al. [6] presented a hybrid model based on EMD and Bi-LSTM–Transformer for predicting the RUL of rolling bearings. EMD was used to decompose the raw vibration signals and extract useful features. Bi-LSTM was used to learn local changes in the degradation process while the Transformer was used to learn long-term relationships in the data. The study used the PHM2012 and XJTU-SY bearing datasets. For data preparation the authors applied EMD to the vibration signals and extracted statistical features from the decomposed signals. The features were then reduced and normalized before being used as model inputs. The proposed model was compared with

(TCN), GRU, and Bi-LSTM. Results showed that the proposed model achieved an average MAE of 0.0469 and RMSE of 0.0563 on PHM2012. On XJTU-SY, it achieved an average MAE of 0.0374 and RMSE of 0.0442.

Wang et al. [7] presented a deep learning method for predicting the RUL of rolling bearings. The method uses a One Dimensional Deep Convolutional Autoencoder (1D-DCAE) to extract useful health features from vibration signals. These features are used to create health indicators that represent the degradation condition of the bearing. The model then uses a Bi-LSTM network with a Temporal Pattern Attention (TPA) mechanism to learn how the bearing condition changes over time. The study used the PHM2012 bearing dataset collected from the PRONOSTIA platform. The dataset contains vibration and temperature signals collected from bearings during accelerated life tests until failure. The vibration signals were sampled at 25.6 kHz for 0.1 seconds every 10 seconds. The authors used the health indicators extracted by 1D-DCAE as labels instead of depending only on traditional linear or piecewise RUL labels. The proposed method was compared with linear labeling and piecewise labeling methods. The authors also tested different versions of the model such as removing the TPA mechanism and using LSTM instead of the multilevel Bi-LSTM. Results showed that the proposed method achieved an average MAE of 0.115 and an average RMSE of 0.079 across seven bearings. It also achieved a Score of 0.699 and performed better than the other labeling methods.

Xiao and his team [8] proposed a model called Multi-Scale Dilated Fusion Attention (MDFA) for predicting RUL. The paper tried to solve some problems related to different operating conditions, multiple degradation patterns and long-term dependencies in sensor data. The study used NASA C-MAPSS, N-CMAPSS and PHM2012 datasets to test the MDFA model. N-CMAPSS is a newer turbofan engine degradation dataset that contains more detailed operating conditions and sensor information. The authors prepared the data by creating RUL labels, using a sliding window technique and applying PCA to select features, reduce dimensionality and convert sensor data into time-series inputs for training. The MDFA model uses Multi-Scale Dilated Convolution to extract degradation features from multiple time scales. It also uses Channel Attention and Spatial Attention mechanisms. These mechanisms help the model focus on important sensors and time points while reducing noise. The model

was compared with CNN, DCNN, TCN, Time-Aware Fully Convolutional Network (TaFCN), Squeeze-and-Excitation Network (SeNet), GRU, CNN-GRU-Attention, and Transformer. The results showed that the MDFA model achieved the best performance in predicting RUL. On C-MAPSS and N-CMAPSS datasets, it achieved MAE values between 0.018 and 0.026, RMSE values between 0.021 and 0.032 and R^2 values higher than 0.987. On PHM2012, it achieved MAE values between 0.023 and 0.026, RMSE values between 0.031 and 0.032 and R^2 values between 0.987 and 0.995.

Özcan [9] studied a maintenance framework for aircraft engines. It combines RUL prediction with dynamic maintenance scheduling. The researcher used the NASA C-MAPSS dataset, which comprises all subsets—FD001, FD002, FD003 and FD004. These refer to varying operating conditions and failure modes. However, the paper focused on a problem traditional maintenance operations employ fixed intervals between service periods. They are not based on real-time sensor monitoring data to overcome this problem. Özcan employed machine learning models, including LightGBM, CatBoost, Gradient Boosting, XGBoost, SVM, KNN and Linear Regression. Grid Search Cross-Validation (Grid Search CV) was used to find the best hyperparameter combinations. The predictions were combined in a framework, ensuring accuracy and robustness. A dynamic maintenance scheduling algorithm was developed based on the RUL threshold. It activates maintenance when the predicted remaining life falls below 15 cycles. The SHapley Additive exPlanations (SHAP) method was employed to understand the models. It highlights the key operational cycles and sensors that influence scheduling decisions. The LightGBM, CatBoost models and their combination observed an RMSE of 6.62 on FD001. The LightGBM, CatBoost and Gradient Boosting combination observed an RMSE of 9.71 on FD003. The other examined cases proved likewise competitive on FD002 and FD004. This confirms that the joint application of RUL estimation with scheduling improves maintenance planning, cuts operating costs and limits risks.

Lei et al [10] proposed an RUL prediction method for rolling bearings based on hierarchical clustering and a Transformer GRU hybrid model. The study used the PHM2012 rolling bearing dataset which contains vibration signals collected from the Pronostia platform under different operating conditions until failure. The preprocessing procedure involved wavelet denoising extracting time domain frequency domain and time frequency domain features smoothing and normalizing the features calculating a comprehensive feature index based on monotonicity and correlation and then removing redundant features using hierarchical clustering with the elbow method within the proposed architecture the Transformer encoder was used to capture dependencies across different time steps and generate high level feature representations while the GRU component was used to model the dynamic degradation trend over time. The proposed approach was compared with GRU TCN Transformer and previous related methods. The results showed that the Transformer GRU model achieved the best predictive performance with average MAE 0.0836 and average RMSE 0.1137 while feature reduction reduced training time by 8.61 percent and model parameters by 23.26 percent.

Jin et al [11] proposed an RUL prediction method for rolling bearings based on TCN-Transformer networks using vibration signals. The study used the IEEE PHM2012 Data Challenge dataset collected from the PRONOSTIA platform. The dataset contains accelerated degradation data from 17 rolling bearings under three operating conditions. The study used horizontal vibration signals because they contain more useful degradation information than vertical signals. The preprocessing steps included extracting 32 time-domain, frequency domain and time-frequency-domain features from the vibration signals. Eight sensitive features were selected using monotonicity, correlation, predictability and robustness. PCA was then used for dimensionality reduction. The authors also constructed a Health Index and smoothed it using a moving average. The proposed TCN-Transformer model used TCN to capture local temporal features and Transformer to capture global relationships. A multi-head feature fusion attention module was then used to combine the local and global features for RUL prediction. The model was compared with RNN, LSTM, GRU, Dual-LSTM, Long Short-Term Memory with Attention Ordered Neurons (LSTM-AON), Bidirectional Gated Recurrent Unit with a gate attention mechanism (BiGRU-GSA), Temporal Convolutional Network with Residual Self-Attention (TCN-RSA) and

Temporal Flow Transformer (TFT). The results showed that TCN-Transformer achieved the best average performance with an RMSE of 0.0514, MAE of 0.0392 and a score of 0.6346. The model reduced RMSE by 14.62% and MAE by 9.26% and improved a score by 13.04% compared with the best previous model.

Chen et al [12] presented a model called Spatial Attention based Convolutional Transformer (SAConvFormer) for predicting the RUL of rolling bearings. The model uses CNN with a spatial attention mechanism to extract local and global features from raw vibration signals. The extracted features are then passed to a Transformer to learn about the long-term relationships related to bearing degradation. The study used the XJTU-SY bearing dataset. The data preparation included determining the First Prediction Time (FPT), normalizing the vibration data, applying a sliding window and dividing the data into training and testing sets. The model directly used raw vibration signals without additional feature engineering. SAConvFormer was compared with Bi-LSTM, Dilated CNN, Multi-Scale Convolutional Neural Network (MSCNN) and Deep Convolutional Neural Network–Multi-Layer Perceptron (DCNN-MLP). The results showed that SAConvFormer achieved strong performance across the bearing datasets. The best result was achieved on Bearing 2_2 with an RMSE of 0.065 and MAE of 0.051. The combination of spatial attention and Transformer also reduced RMSE by 23.7% and MAE by 16.9% in the Bearing 3_1 experiment.

Zou et al. [13] proposed an RUL prediction method based on an Attention-based Double Bidirectional Long Short Term Memory network (A-DBiLSTM). The study used the PHM2012 and XJTU-SY bearing datasets. For data preparation, Discrete Wavelet Transform (DWT) was used to reduce noise from vibration signals. Features were then extracted from the time frequency, and time frequency domains. The most useful degradation features were selected using monotonicity, predictability, trendability and robustness. Kernel Principal Component Analysis (KPCA) was then used to combine the selected features and create a Health Index (HI). The 3-sigma

method was also used to determine the bearing failure threshold. The A-DBiLSTM model used two Bi-LSTM layers with an attention mechanism to learn the degradation pattern and focus on important information. Bayesian optimization was used to select the model hyperparameters. The proposed model was compared with Bidirectional Gated Recurrent Unit (BiGRU) and CNN-LSTM. Results showed that A-DBiLSTM achieved lower RMSE and MAE values and higher R^2 values than the comparison models on both PHM2012 and XJTU-SY. The statistical tests also showed significant differences with p-values below 0.05, which supported the reliability of the prediction results.

Yang et al. [14] proposed a deep multiscale feature fusion network with dual attention for predicting the RUL of rolling bearings. The model first uses a Squeeze and Excitation attention mechanism (SE) to give more importance to useful time-step information. A multi-scale residual pyramid module is then used to extract and combine degradation features at different scales. The model also uses a Transformer encoder with enhanced multi-head attention and relative position encoding to learn relationships between different time steps. The study used the XJTU-SY bearing dataset. The authors extracted features from the time and frequency domains and normalized them before using them as model inputs. The proposed model was compared with Transformer Encoder, SE Transformer Encoder, Gated Convolutional Unit Transformer (GCU Transformer), Temporal Convolutional Network with Multi-head Attention (TCN-MA) and Informer-Encoder. The results showed that the proposed model achieved the best prediction performance. On Bearing2_2, it achieved an RMSE of 0.05301, MAE of 0.04159 and R^2 of 0.96186. On Bearing2_5, it achieved an RMSE of 0.07505, MAE of 0.06182 and R^2 of 0.93149.

Liu et al. [15] proposed an RUL prediction model for rolling bearings called TcLstmNet with a Convolutional Block Attention Module (TcLstmNet CBAM). The model combines TCN and LSTM in parallel. TCN is used to learn long term temporal relationships and deeper spatial features while LSTM is used to learn short term temporal relationships. The extracted features are then combined and passed through the Convolutional Block Attention Module (CBAM), which gives more importance to

useful features. A fully connected layer is used at the end to predict RUL. The study used the PHM2012 and XJTU-SY bearing datasets. The model used raw vibration signals as input. The data was divided into training and testing sets, and horizontal and vertical vibration signals were extracted before feature learning. The proposed model was compared with RNN, TCN, LSTM, Soft-Threshold Attention Mechanism-based Temporal Convolutional Network (TCN-SA) and Temporal Convolutional Network with Residual Separable Convolutional Block (TCN-R). Results showed that TcLstmNet-CBAM achieved better performance than the comparison models. The model achieved an MAE of 2.287 and an RMSE of 3.123.

Table 2.4. Comparison of Related Studies

P a p e r	All Mo d e l s	D a t a s e t s	Dataset Link	Limit a t i o n	Best Model	Result s
D e n g a n d Z h o u [1]	CN N, LST M, CN N- LST M, BP, GR U, CN N- GR U, DC NN, TC N	N A S A C - M A P S S F D O O 1 , F D O O 2 , F D O O 3 ,	https://data.nasa.gov/dataset/cm-apss-jet-engine-simulated-data	The stud y focu sed main ly onR ULpr edict ion and was teste d only on simu lated turb ofan engi ne data.	CNN- LSTM with attentio n mechani sm	RMSE: FD001 = 15.97 7, FD002 = 14.45 2, FD003 = 13.90 7, FD004 = 16.63 7

		FD004				
F a n e t a l. [2]	STAR, CNN-based models, RN-based models, DA-ST, DLF or mer, BiLSTM-DAE-Transformer	NASCMAPSFD001, FD002, FD003, FD004	https://data.nasa.gov/dataset/cm-apss-jet-engine-simulated-data	The hierarchical Transformer may increase computational complexity and was tested on simulated engine data.	STAR	RMSE: FD001 = 10.61, FD002 = 13.47, FD003 = 10.71, FD004 = 15.87. Score: 169, 784, 202 and 1449
E l i s h e r i f e t a l. [3]	CAELSTM and other RUL prediction approaches	NASCMAPSFD001, FD002, FD003, FD004	https://data.nasa.gov/dataset/cm-apss-jet-engine-simulated-data	The model was tested only on FD001 and FD003 and needs validation on more complex oper	CAELSTM	RMSE: FD001 = 14.44, FD003 = 13.40. MAE: FD001 = 10.49, FD003 = 10.68. Score: 282.38 and 264.47

		03		ating conditions.		
E n s a r i o g l u e t a l . [4]	1D CNN-LSTM, 1D CNN-LSTM with difference feature, traditional machine learning and deep learning models	N A S A C - M A P S S F D O O 1	https://data.nasa.gov/dataset/cm-apss-jet-engine-simulated-data	The method was tested only on FDO1 which has one operating condition and one fault mode.	1D CNN-LSTM with difference feature and change point based labeling	RMSE = 16.1, Score = 437.2
S u n e t a l . [5]	MD SCT, TCN-SA, DANN, TCN-RSCB, CVT, CLSTM, MCNN, SAGC, NN	P H M 2 0 1 2 a n d X J T U - S Y	PHM2012: https://github.com/wkzs111/phm-ieee-2012-data-challenge-dataset XJTU-SY: https://biaowan.g.tech/xjtu-sy-bearing-datasets/	The model has a complex architecture and its performance varies between different bearings	MDSCT	PHM2012: MAE = 0.100, RMSE = 0.124. XJTU-SY: MAE = 0.134, RMSE = 0.160, R ² = 0.667

	SA, DC NN			cond ition s.		
Ji n e t a l. [6]	EM D- BiL ST M- Tra nsf or me r, TC N, GR U, Bi- LST M	P H M 2 0 1 2 a n d X J T U - S Y	PHM2012: https://github.com/wkzsl111/phm-ieee-2012-data-challenge-dataset XJTU-SY: https://biaowan.g.tech/xjtu-sy-bearing-datasets/	The met hod requ ires sign al deco mpo sitio n and feat ure proc essin g befo re train ing.	EMD- BiLSTM- Transfor mer	PHM2 012: MAE = 0.046 9, RMSE = 0.056 3. XJTU- SY: MAE = 0.037 4, RMSE = 0.044 2
W a n g e t a l. [7]	1D- DC AE, Bi- LST M, Bi- LST M wit h TP A, line ar lab elin g and pie ce wis e lab elin g me tho ds	P H M 2 0 1 2	https://github.com/wkzsl111/phm-ieee-2012-data-challenge-dataset	The stud y was eval uate d on a limit ed num ber of beari ngs and depe nds on learn ed healt h indic ators .	1D-DCAE with Bi- LSTM and TPA	Averag e MAE = 0.115, averag e RMSE = 0.079, Score = 0.699
X i a o e t	CN N, DC NN, TC N,	N A S A C -	C-MAPSS: https://data.nasa.gov/dataset/cmmapss-jet-engine-	The mod el uses mult iple	MDFA	C- MAPS S and N- CMAP SS:

a l. [8]	TaF CN, Se Net , GR U, CN N- GR U- Att ent ion, Tra nsf or me r, MD FA	M A P S S , N - C M A P S S a n d P H M 2 0 1 2	simulated-data PHM2012: https://github.com/wkzs111/phm-ieee-2012-data-challenge-dataset	conv oluti on and atte ntio n mod ules whic h may incre ase com puta tion al com plexi ty.		MAE = 0.018 - 0.026, RMSE = 0.021 - 0.032, R ² > 0.987. PHM2 012: MAE = 0.023 - 0.026, RMSE = 0.031 - 0.032, R ² = 0.987 - 0.995
O z k a n [9]	Lig htG BM , Cat Bo ost, Gra die nt Bo osti ng, XG Bo ost, SV M, KN N, Lin ear Reg res sio n and ens em ble mo del s	N A S A C - M A P S S F D 0 0 1 , F D 0 0 2 , F D 0 0 3 , F D 0 0 4	https://data.nasa.gov/dataset/cm-apss-jet-engine-simulated-data	The fram ewo rk was teste d only on C- MAP SS and used a fixed main tena nce thre shol d of 15 cycle s.	LightGB Mand CatBoost baseden sembles	FD001 RMSE = 6.62. FD003 RMSE = 9.71. FD002 RMSE = 10.15. FD004 RMSE = 14.12
L e i	GR U, TC	P H M	https://github.com/wkzs111/phm-ieee-2012-data-challenge-dataset	The met hod	Transfor mer- GRU	Averag e MAE =

e t a l. [1 0]	N, Tra nsf or me r, Tra nsf or me r- GR U	2 0 1 2	m-ieee-2012- data-challenge- dataset	requ ires seve ral feat ure extr action and selec tion steps befo re mod el train ing.	with hierarchi cal clusterin g	0.083 6,aver ageR MSE = 0.113 7. Traini ng time reduc ed by 8.61% and model param eters by 23.26 %
Ji n e t a l. [1 1]	RN N, LST M, GR U, Du al- LST M, LST M- AO N, BiG RU- GS A, TC N- RS A, TFT / TC N- Tra nsf or me r	I E E E P H M 2 0 1 2 D a t a C h a l l e n g e	https://github.co m/wkzs111/ph m-ieee-2012- data-challenge- dataset	The mod el still has large r error s duri ng later degr adati on stag es beca use failu re beha vior can beco me highl y nonli near .	TCN- Transfor mer	RMSE = 0.051 4, MAE = 0.039 2, Score = 0.634 6. RMSE reduc ed by 14.62 %, MAE by 9.26% and Score impro ved by 13.04 %
C h e n e t a l. [1 1]	SAC onv For me r, Bi- LST M, Dila ted CN N,	X J T U - S Y	https://biaowan g.tech/xjtu-sy- bearing- datasets/	The mod el was teste d main ly on one beari ng data	SAConvF ormer	Bearin g2_2: RMSE = 0.065, MAE = 0.051. On Bearin g3_1, RMSE reduc

2]	MS CN, DC NN - ML P			set and need s valid atio n on mor e indu strial cond ition s.		ed by 23.7% and MAE by 16.9%
Z o u e t a l. [1 3]	A- DBi LST M, BiG RU, CN N- LST M	P H M 2 0 1 2 a n d X J T U - S Y	PHM2012: https://github.com/wkzs111/phm-ieee-2012-data-challenge-dataset XJTU-SY: https://biaowan.g.tech/xjtu-sy-bearing-datasets/	The met hod requ ires seve ral prep roce ssing step s inclu ding DWT , feat ure extr actio n, KPC Aan dBay esia n opti miza tion.	A- DBiLSTM	The model achiev ed lower RMSE and MAE and higher R ² than the compa rison model s. Statist ical tests report ed p- values below 0.05
Y a n g e t a l. [1 4]	Tra nsf or me r Enc ode r, SE Tra nsf or me r Enc ode r, GC U Tra	X J T U - S Y	https://biaowan.g.tech/xjtu-sy-bearing-datasets/	The mod el uses seve ral feat ure extr actio n and atte ntio n mod ules whic h may	Deep multiscal e feature fusion network with dual attentio n	Bearin g2_2: RMSE = 0.053 01, MAE = 0.041 59, R ² = 0.961 86. Bearin g2_5: RMSE = 0.075 05, MAE = 0.061

	nsf or me r, TC N- MA , Inf or me r- Enc ode r, pro pos ed dua l- att ent ion mo del			incre ase mod el com plexi ty.		82, R^2 = 0.931 49
L i u e t a l. [1 5]	TcL stm Net - CB AM , RN N, TC N, LST M, TC N- SA, TC N-R	P H M 2 0 1 2 a n d X J T U - S Y	PHM2012: https://github.com/wkzs111/phm-ieee-2012-data-challenge-dataset XJTU-SY: https://biaowan.g.tech/xjtu-sy-bearing-datasets/	The mod el com bine s seve ral paral lel and atte ntio n mod ules whic h may incre ase com puta tion al requ irem ents.	TcLstmN et-CBAM	MAE = 2.287, RMSE = 3.123

Chapter 3: Proposed Approach (Methodology)

Chapter 5: Conclusion and Future Work

References

References

- [1] Deng, S., and Zhou, J. "Prediction of Remaining Useful Life of Aero-engines Based on CNN-LSTM-Attention." *International Journal of Computational Intelligence Systems*, vol. 17, article 232, 2024. DOI: 10.1007/s44196-024-00639-w.
- [2] Fan, Z., Li, W., and Chang, K.-C. "A Two-Stage Attention-Based Hierarchical Transformer for Turbofan Engine Remaining Useful Life Prediction." *Sensors*, vol. 24, no. 3, article 824, 2024. DOI: 10.3390/s24030824.
- [3] Elsherif, S. M., Hafiz, B., Makhlouf, M. A., and Farouk, O. "A Deep Learning-Based Prognostic Approach for Predicting Turbofan Engine Degradation and Remaining Useful Life." *Scientific Reports*, vol. 15, article 26251, 2025. DOI: 10.1038/s41598-025-09155-z.

- [4] Ensarioğlu, K., İnkaya, T., and Emel, E. “Remaining Useful Life Estimation of Turbofan Engines with Deep Learning Using Change-Point Detection Based Labeling and Feature Engineering.” *Applied Sciences*, vol. 13, no. 21, article 11893, 2023. DOI: 10.3390/app132111893.
- [5] Sun, N., Tang, J., Ye, X., Zhang, C., Zhu, S., Wang, S., and Sun, Y. “Remaining Useful Life Prognostics of Bearings Based on Convolution Attention Networks and Enhanced Transformer.” *Heliyon*, vol. 10, no. 19, article e38317, 2024. DOI: 10.1016/j.heliyon.2024.e38317.
- [6] Jin, C., Li, B., Yang, Y., Yuan, X., Tu, R., Qiu, L., and Chen, X. “Remaining Useful Life Prediction of Rolling Bearings Based on Empirical Mode Decomposition and Transformer Bi-LSTM Network.” *Applied Sciences*, vol. 15, no. 17, article 9529, 2025. DOI: 10.3390/app15179529.
- [7] Wang, C., Jiang, W., Shi, L., et al. “Rolling Bearing Remaining Useful Life Prediction Using Deep Learning Based on High-Quality Representation.” *Scientific Reports*, vol. 15, article 8228, 2025. DOI: 10.1038/s41598-025-93165-4.
- [8] Xiao, G., Jin, C., and Bai, J. “Remaining Useful Life Prediction for Aero-Engines Based on Multi-Scale Dilated Fusion Attention Model.” *Applied Sciences*, vol. 15, no. 17, article 9813, 2025. DOI: 10.3390/app15179813.
- [9] Özcan, H. “Interpretable Ensemble Remaining Useful Life Prediction Enables Dynamic Maintenance Scheduling for Aircraft Engines.” *Scientific Reports*, vol. 15, article 39795, 2025. DOI: 10.1038/s41598-025-23473-2.
- [10] Lei, W., Dong, X., Cui, F., and Huang, G. “A Remaining Useful Life Prediction Method for Rolling Bearings Based on Hierarchical Clustering and Transformer-GRU.” *Applied Sciences*, vol. 15, no. 10, article 5369, 2025. DOI: 10.3390/app15105369.
- [11] Jin, X., Ji, Y., Li, S., Lv, K., Xu, J., Jiang, H., and Fu, S. “Remaining Useful Life Prediction for Rolling Bearings Based on TCN-Transformer Networks Using Vibration Signals.” *Sensors*, vol. 25, no. 11, article 3571, 2025. DOI: 10.3390/s25113571.
- [12] Chen, C., Wang, T., Liu, Y., Cheng, L., and Qin, J. “Spatial-Attention-Based Convolutional Transformer for Bearing Remaining Useful Life Prediction.” *Measurement Science and Technology*, vol. 33, no. 11, article 114001, 2022. DOI: 10.1088/1361-6501/ac7c5b.
- [13] Zou, Y., Sun, W., Wang, H., Xu, T., and Wang, B. “Research on Bearing Remaining Useful Life Prediction Method Based on Double Bidirectional Long Short-Term Memory.” *Applied Sciences*, vol. 15, no. 8, article 4441, 2025. DOI: 10.3390/app15084441.
- [14] Yang, Y., Wang, Z., Liu, X., Liu, T., et al. “Deep Multiscale Feature Fusion Network with Dual Attention for Rolling Bearing Remaining Useful Life Prediction.” *Scientific Reports*, vol. 15, article 14868, 2025. DOI: 10.1038/s41598-025-97380-x.
- [15] Liu, Q., Dai, Z., Lai, H., Chen, M., Huang, H., Fu, J., Hou, M., Xu, X., Wang, G., et al. “A Novel RUL Prediction Method for Rolling Bearing: TcLstmNet-CBAM.” *Scientific Reports*, vol. 15, article 14055, 2025. DOI: 10.1038/s41598-025-98845-9.

- [16] Scikit-learn Developers. "LinearRegression." Scikit-learn Documentation, accessed August 6, 2026.
- [17] Scikit-learn Developers. "Support Vector Regression." Scikit-learn Documentation, accessed August 6, 2026.
- [18] Scikit-learn Developers. "Nearest Neighbors Regression." Scikit-learn Documentation, accessed August 6, 2026.
- [19] Scikit-learn Developers. "RandomForestRegressor." Scikit-learn Documentation, accessed August 6, 2026.
- [20] Scikit-learn Developers. "GradientBoostingRegressor." Scikit-learn Documentation, accessed August 6, 2026.
- [21] Microsoft Corporation. "LightGBM Documentation." LightGBM, accessed August 6, 2026.
- [22] CatBoost Developers. "CatBoost Documentation." CatBoost, accessed August 6, 2026.
- [23] XGBoost Developers. "XGBoost Documentation." XGBoost, accessed August 6, 2026.
- [24] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. DOI: 10.1109/5.726791.
- [25] Hochreiter, S., and Schmidhuber, J. "Long Short-Term Memory." *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.
- [26] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. "Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, 2014.
- [27] Graves, A., and Schmidhuber, J. "Framewise Phoneme Classification with Bidirectional LSTM Networks." *Proceedings of the International Joint Conference on Neural Networks*, pp. 2047–2052, 2005. DOI: 10.1109/IJCNN.2005.1556215.
- [28] Hinton, G. E., and Salakhutdinov, R. R. "Reducing the Dimensionality of Data with Neural Networks." *Science*, vol. 313, no. 5786, pp. 504–507, 2006. DOI: 10.1126/science.1127647.
- [29] Bahdanau, D., Cho, K., and Bengio, Y. "Neural Machine Translation by Jointly Learning to Align and Translate." *International Conference on Learning Representations*, 2015.
- [30] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. "Attention Is All You Need." *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

- [31] Bai, S., Kolter, J. Z., and Koltun, V. “An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling.” *arXiv preprint arXiv:1803.01271*, 2018.
- [32] Scikit-learn Developers. “Root Mean Squared Error.” *Scikit-learn Documentation*, accessed August 6, 2026.
- [33] Scikit-learn Developers. “Mean Absolute Error.” *Scikit-learn Documentation*, accessed August 6, 2026.
- [34] Scikit-learn Developers. “R² Score.” *Scikit-learn Documentation*, accessed August 6, 2026.
- [35] Saxena, A., Goebel, K., Simon, D., and Eklund, N. “Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation.” *2008 International Conference on Prognostics and Health Management*, pp. 1–9, 2008. DOI: 10.1109/PHM.2008.4711414.