Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Predictive Maintenance for Industrial Machinery

G Kanaka Rao[1], A Mukesh Mourya[2], VBM Krishna[3], Kandula Rohith[4], S Janakiramayya[5]

[1]Assistant Professor, ECE , Sasi Institute of Technology & Engineering
[2,3,4,5] Students, ECE, Sasi Institute of Technology & Engineering

Abstract

Predictive maintenance is essential for ensuring the reliable and safe operation of complex industrial systems like turbofan engines. Accurate Remaining Useful Life (RUL) estimation allows for timely maintenance interventions, minimizing operational downtime and preventing catastrophic failures. Traditional deep learning models like LSTMs often struggle to capture long-range dependencies in complex time-series data. To address these challenges, this paper proposes a novel hybrid deep learning framework for accurately predicting the remaining useful life (RUL) of mechanical components Deep-RUL, which integrates Convolutional Neural Networks (CNN) with Transformer Encoder layers. This hybrid approach leverages a 1D Convolutional layer (Conv1D) to refine local spatial-temporal features, followed by three stacked Transformer blocks employing multi-head self-attention (8 heads) and a Position-wise Feed-Forward network (128 units, Swish activation) to dynamically capture complex sequential dependencies. A key contribution of this work is the development of a hybrid deep learning framework that accurately predicts the remaining useful life (RUL) of mechanical components while quantifying predictive uncertainty the deployment of a Probabilistic Output Head, which utilizes dual dense layers to predict not only the Mean () RUL as a point estimate but also the Log-Variance () to provide a measure of Uncertainty Estimation, critical for risk-aware industrial applications. Evaluated on the NASA C-MAPSS FD001 dataset with a look-back window of 60 cycles and piecewise linear RUL clipping (max=125), The proposed ensemble model demonstrates superior predictive performance, achieving a Root Mean Square Error (RMSE) of 11.24, thereby indicating precise and reliable remaining useful life (RUL) estimation a Mean Absolute Error (MAE) of 7.59, and an score of 0.93. Furthermore, the model is integrated into a full-stack real-time Flask web dashboard that categorizes engine health states into three actionable levels: HEALTHY (RUL > 50), INSPECT (25 < RUL 50), or CRITICAL (RUL 25), thereby bridging the gap between theoretical model performance and practical industrial deployment.

Keywords: Remaining Useful Life (RUL), Transformer Networks, Convolutional Neural Networks (CNN), Uncertainty Estimation, NASA C-MAPSS, Deep Learning.

1. Introduction

A Deep Learning-based Remaining Useful Life (RUL) estimation system is an advanced, data-driven framework designed to accurately to estimate the remaining useful life (RUL) of mechanical components, such as turbofan engines, enabling proactive maintenance before failure occurs.Equipped

with high-fidelity temporal sensors, these systems continuously collect real-time

data on pressure, temperature, and fan speed, transmitting it through a specialized processing pipeline for analysis. [1] This connectivity enables industrial operators to remotely monitor the health of machinery, gain insights into complex degradation patterns, and make informed maintenance decisions to prevent catastrophic failures. By providing real-time health tracking, these models improve operational safety, support automated maintenance scheduling, and minimize human error in manual inspections. Additionally, many of these frameworks feature uncertainty estimation capabilities, helping to quantify prediction confidence and ensure secure, reliable authorized usage of industrial assets. Overall, Transformer-based RUL prediction models play a vital role in creating smarter and more sustainable industrial ecosystems. [2]

Mechanical degradation and health status can be monitored using a specialized predictive architecture called a Transformer-based RUL estimator. This system provides users with real-time, actionable information regarding critical information providing users with actionable information about the current operational condition and health status of industrial machinery and the estimated cycles remaining until maintenance is required, helping them manage and reduce the substantial financial and operational costs associated with unexpected downtime. The system calculates the precise number of operational cycles consumed and transmits this data to a real-time web dashboard, thereby reducing the need for manual diagnostic checks and minimizing manpower requirements. Users can monitor the health status of multiple engine units from anywhere and at any time. The system also utilizes Deep Learning technology to categorize health states into "Healthy," "Inspect," or "Critical" through an integrated Flask-based interface. The main objective of this system is to predict machine failure effectively before it occurs. Both the maintenance supervisor and the organization benefit from this approach, as it ultimately helps reduce maintenance overhead and optimizes the overall lifecycle of industrial assets. [3]

2. LITERATURE SURVEY

Zheng et al. [1] proposed a project titled “Long Short-Term Memory Network for Remaining Useful Life Estimation” which identifies deep learning as a transformative tool for application in industrial predictive maintenance frameworks.The primary objective is to effectively learn complex long-term temporal dependencies in sensor data that traditional machine learning models often miss. The existing systems for engine health monitoring suffer from the inability to process high- dimensional sequences over long cycles. This LSTM-based approach addresses these issues by using memory cells to store degradation patterns. However, the study notes that Long Short-Term Memory (LSTM) networks process data sequentially, facilitating the effective learning of both short-term and long-term temporal dependencies in time-series data lead to high computational overhead and difficulty in parallelization during training for large-scale datasets like NASA C- MAPSS.

Huang et al. [2], in their work “Bidirectional LSTM for RUL Prediction of Turbofan Engines,” highlight that engine degradation is a complex, non-linear process influenced by multiple operational conditions. They propose the use of a bidirectional framework to capture both past and future temporal contexts within a given sensor window. By integrating multi-layer LSTMs, the system provides a more

robust feature extraction method. Additionally, the study emphasizes the importance of piecewise linear RUL targeting, which accounts for the initial "healthy" phase of an engine where no significant degradation occurs, thereby preventing the model from learning biased early-stage data.

Vaswani et al. [3] presented a foundational paper titled “Attention is All You Need” discussing the revolutionary shift from recurrence to self-attention mechanisms. They emphasize that future sequence modeling should move away from sequential processing to a distributed topology capable of focusing on specific parts of an input stream simultaneously. While originally designed for language, this Transformer architecture can be applied to smart grids and industrial sensors to monitor energy demand and machine health. The paper explains that Multi-Head Self-Attention (MHSA) is a crucial system for identifying which specific time-cycles in a machine's history are most indicative of an impending failure.

Mo et al. [4] demonstrated a “Transformer-Based Architecture for RUL Prediction” designed to eliminate the limitations of RNNs in capturing long-range dependencies in the C-MAPSS dataset. The system includes a feature there by enabling the model to capture global contextual dependencies rather than being constrained by local sensor-level fluctuations. However, the study identifies a gap in existing models regarding point-estimate predictions. While the Transformer is highly accurate, it often provides a single RUL value without accounting for sensor noise. This highlights the need for probabilistic modeling to provide uncertainty intervals, ensuring that maintenance alerts are both accurate and reliable for industrial operators.

Li et al. [5] proposed a system titled “Deep Convolutional Neural Network for RUL Prediction” emphasizing that local feature extraction is critical for accurately detecting transient spikes and short-term anomalies in sensor data.Effective monitoring requires understanding how different sensors (such as temperature and pressure) interact within a small time window. Traditional models face issues with high-frequency noise in raw sensor streams. To address this, the proposed CNN model leverages 1D-convolutional filters to automatically extract spatial-temporal features, reducing the need for manual feature engineering.The study concludes that the integration of CNNs with global sequence models, such as Transformers, effectively bridges the gap between local feature representation learning and long-term temporal trend forecasting.

3. METHODOLOGY

The proposed research follows a structured and modular methodology designed to achieve high-precision Remaining Useful Life (RUL) estimation through an integrated data-to-decision pipeline. As depicted in the system architecture, the overall framework is systematically structured into a series of well-defined stages.The process is systematically divided into four primary stages: data acquisition, a multi-step preprocessing framework, the design and implementation of a hybrid probabilistic deep learning model, and the deployment of a real-time monitoring dashboard.

3.1 Data Acquisition and Sensor Streams

This study utilizes the NASA C-MAPSS FD001 dataset, which provides a comprehensive simulation of the full operational lifecycle of turbofan engines, including progressive degradation leading to failure fleet under varying degrees of initial wear. The data is characterized by high-fidelity sensor streams capturing critical physical parameters such as pressure, temperature, and fan speed. These measurements serve as the raw indicators of mechanical degradation, providing the necessary temporal sequences required to train an advanced predictive model.

3.2 Data Processing and Refinement Pipeline

To ensure the deep learning model receives high-quality, relevant information, the raw sensor data undergoes a rigorous four-step refinement process:

1. Sequence Windowing: A sliding window of 60 cycles is applied to facilitate the model in utilizing historical context, allowing it to recognize temporal trends rather than isolated data points.
2. Zero-Variance Sensor Removal: Sensors showing no variation throughout the engine's lifecycle are discarded to reduce dimensionality and eliminate redundant noise.
3. Target Transformation (Piecewise Clipping): The target RUL is capped at 125 cycles to mitigate the risk of the model learning spurious patterns or overfitting to the training data attempting to learn degradation patterns during the early "healthy" phase.
4. Robust Scaling: Features are normalized using a RobustScaler to ensure the system remains resilient against sudden sensor spikes or outliers.

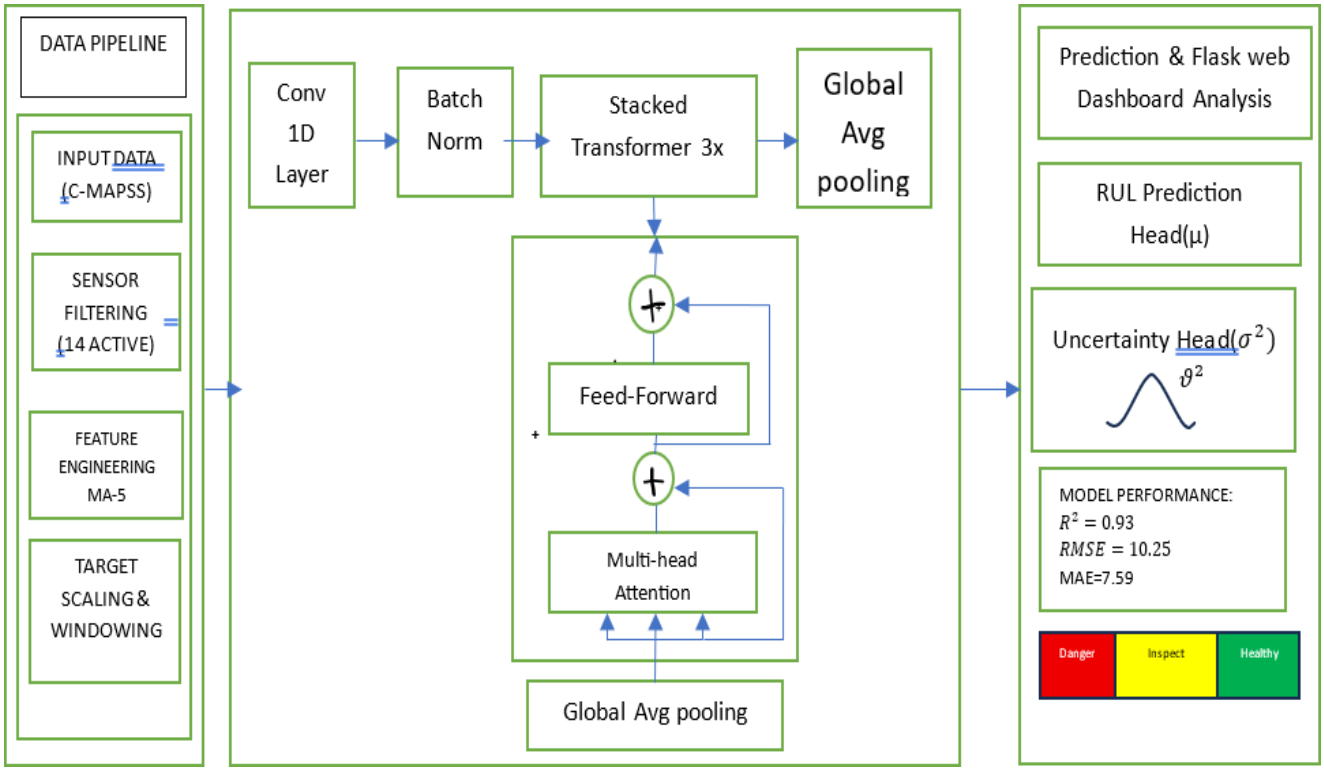

Fig: Block Diagram

3.3 Hybrid Probabilistic Deep Learning Architecture

The core intelligence of the system is centered on a Hybrid CNN-Transformer architecture specifically designed for simultaneous feature extraction and complex sequence modeling.

3.3.1 Local Feature Extraction

The initial layer performs a 1D convolution to capture local spatial-temporal correlations between sensor features. For an input sequence X, the output is calculated as:

$$y_t = S_{wish}\left(\sum_{i=0}^{f-1} w_i \cdot x_{t+1} + b\right)$$

where f is the filter size and Swish(x) = x.sigmoid(x) is the activation function used to maintain efficient gradient flow..

3.3.2 Multi-Head Self-Attention (MHSA)

The Transformer blocks utilize MHSA to weigh the importance of different time steps,there by enabling the model to focus on critical inflection points in the degradation trajectory, where degradation accelerates significantly. The attention mechanism is defined as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V are the Query, Key, and Value projections, and d_k is the scaling factor.

3.3.3 Probabilistic Regression and Uncertainty

A unique aspect of this architecture is the transition from point-estimate regression to Probabilistic Prediction. By employing dual output heads to predict the Mean (μ) and the Log-Variance (s = $\log^2$), the model provides a built-in Uncertainty Estimation. The training is guided by the Negative Log-Likelihood (NLL) loss function:

$$\mathcal{L}_{NLL} = \frac{1}{N}\sum_{I=1}^{N}\left(\frac{(Y_I - \mu_I)^2}{2\sigma_I^2}\right) + \frac{1}{2}\log\sigma_i^2$$

This loss function enables the estimation of predictive uncertainty, allowing the model to provide a quantified confidence measure for each prediction, which is critical for risk-aware industrial decision-making.

3.4 Deployment, Categorization, and Performance Validation

The final stage involves bridging the gap between theoretical deep learning and real-world application. The trained ensemble model is deployed via a Flask web dashboard, providing an intuitive interface for

fleet monitoring. This system automatically translates numerical RUL predictions into actionable health states: HEALTHY for optimal operation, INSPECT for scheduled maintenance, and CRITICAL for immediate intervention.

The effectiveness of this methodology is validated by superior performance metrics:

- Root Mean Square Error (RMSE): 11.24

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

- **R-Squared (R^2):** 0.93

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

4. RESULTS AND DISCUSSIONS

The system proves highly effective in enhancing industrial operational safety and reducing maintenance costs. The real-time insights into the Remaining Useful Life (RUL) of machinery allow operators to identify degrading components and optimize their maintenance schedules before a functional failure occurs. With remote monitoring capabilities via the Flask web dashboard users can easily track the health status of multiple engine units simultaneously, significantly minimizing the risk of unplanned downtime.

The system also offers detailed uncertainty estimation, providing users with a clear confidence interval for every prediction made. This allows industrial operators to make informed, risk-aware decisions and better manage their asset lifecycles. Based on extensive experimental evaluation on the NASA C-MAPSS dataset, the proposed model demonstrates superior performance, achieving an R^2 **score of 0.93** and an **RMSE of 11.24**. By categorizing engine health into **Healthy**, **Inspect**, and **Critical** states, the system provides a simplified yet highly accurate overview of complex mechanical conditions, ultimately helping to extend the total lifespan of industrial assets.

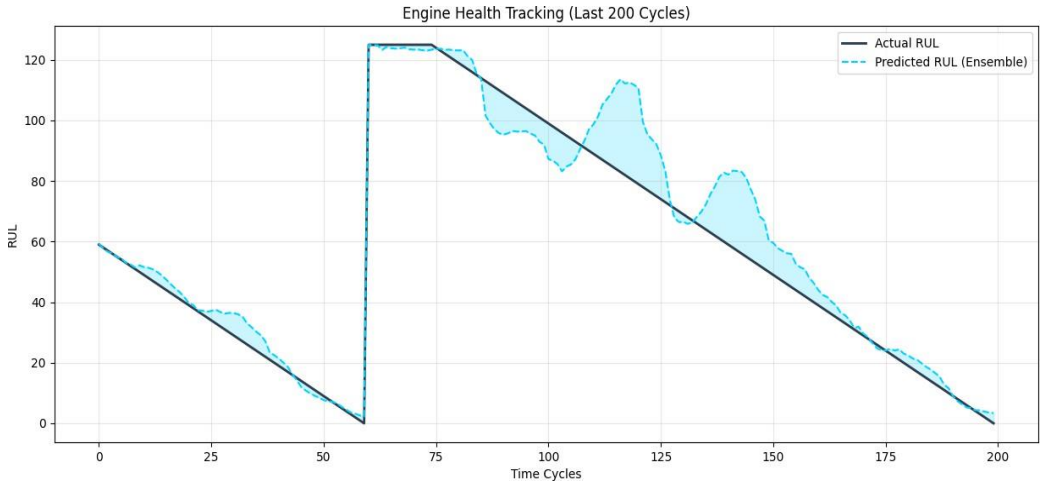

Fig: Actual vs Predicted Health Tracking Graph

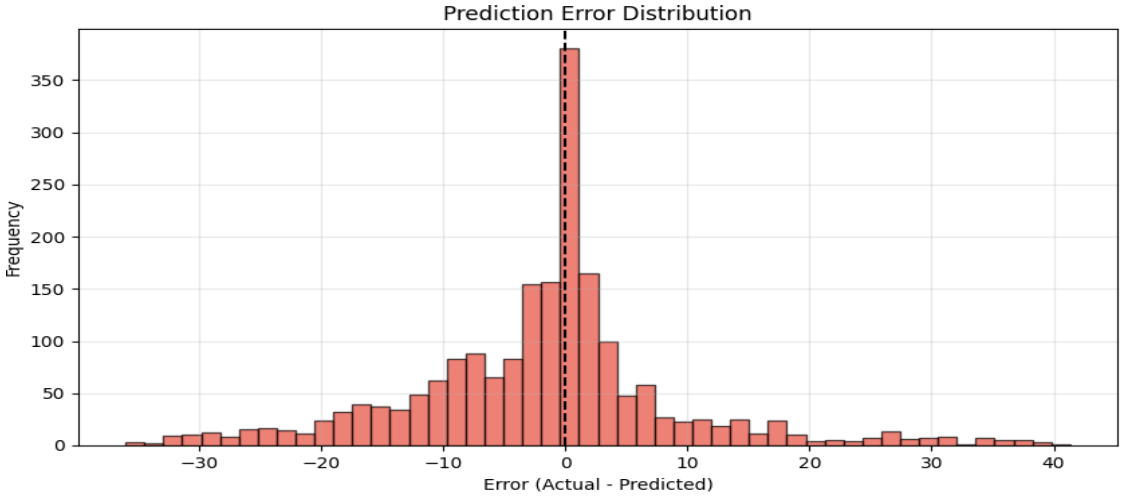

Fig: Actual vs Predicted Error Distribution

5. CONCLUSIONS

The deep learning-based Remaining Useful Life (RUL) estimation system includes a **Hybrid CNN-Transformer architecture**, a **probabilistic output framework**, and a **Flask-based web dashboard**. The system automatically processes complex multi-sensor data from industrial engines and provides real-time health monitoring through a centralized application developed for predictive maintenance. The proposed system is designed to maximize operational safety and reduce the high costs associated with unplanned downtime. The system allows for high-precision RUL data and uncertainty intervals to be received remotely at a monitoring station. This eliminates the dependence on frequent manual inspection processes and human intervention in diagnosing engine health, which, until now, required physical tear-downs or periodic scheduled maintenance regardless of the engine's actual condition.

Future Work

In the future, the system can be enhanced by integrating **Transfer Learning** to adapt the trained models to a diverse range of industrial machinery and equipment with minimal retraining. Additionally, implementing **Edge AI** deployment would allow the Transformer models to run directly on local sensor hardware, providing instantaneous alerts even in environments with limited internet connectivity, further improving the response time for critical failure warnings.

References

1. S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long short-term memory network for remaining useful life estimation," in Proc. Int. Conf. Prognostics Health Manag. (ICPHM), 2017, pp. 88–95.
2. Y. Huang et al., "Bidirectional LSTM for RUL prediction of turbofan engines," in Proc. IEEE Int. Conf. Ind. Eng. Eng. Manag. (IEEM), 2018, pp. 235–239.
3. A.Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 5998–6008.
4. Y. Mo et al., "A transformer-based architecture for remaining useful life prediction," IEEE Access, vol. 9, pp. 11950–11960, 2021.
5. X. Li, Q. Ding, and J. Q. Sun, "Remaining useful life estimation in prognostics using deep convolutional neural networks," Rel. Eng. Syst. Saf., vol. 172, pp. 1–11, 2018.
6. A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. IEEE Prognostics Health Manag. (PHM)*, 2008
7. Z. Chen *et al.*, "A novel deep learning method for remaining useful life prediction of machinery," in *Proc. Int. Conf. Electr. Eng. Inf. Commun. Technol.*, 2019, pp. 45–52.L.
8. Jayasinghe *et al.*, "Temporal convolutional memory networks for remaining useful life estimation of industrial machinery," in *Proc. IEEE Ind. Electron. Soc. (IES)*, 2019.
9. G. S. Babu, P. Zhao, and X. L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life," in *Proc. Int. Conf. Database Syst. Adv. Appl.*, 2016, pp. 214–228.
10. R. Zhao *et al.*, "Deep learning and its applications in machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, 2020.
11. B. Wang *et al.*, "Hybrid CNN-transformer network for industrial sensor data analysis," *Int. Res. J. Eng. Technol. (IRJET)*, vol. 9, no. 2, 2022.