Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

A Risk-Aware Explainable Multi-Agent Decision Intelligence Framework for Human-Centric Predictive Maintenance in Smart Manufacturing

Naman Bafna
MS Ramaiah University of Applied Sciences

Pranav Kumar D
MS Ramaiah University of Applied Sciences

Prathap Rudra Boppuru
boppururudraprathap.cs.et@msruas.ac.in

MS Ramaiah University of Applied Sciences https://orcid.org/0000-0002-5161-4972

Research Article

Keywords: Predictive Maintenance, Remaining Useful Life Prediction, Explainable Artificial Intelligence (XAI), Agentic Artificial Intelligence, Prognostics and Health Management (PHM), Multi-Agent Decision Intelligence, Reliability-Aware Prognostics, Risk-Aware Maintenance Decision Support

Posted Date: October 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10633014/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

A Risk-Aware Explainable Multi-Agent Decision Intelligence Framework for Human-Centric Predictive Maintenance in Smart Manufacturing

Naman Bafna, Pranav Kumar D, Prathap Rudra Boppuru,*
Department of Computer Science and Engineering
M S Ramaiah University of Applied Sciences, Bengaluru, India.

**boppururudraprathap.cs.et@msruas.ac.in*

ABSTRACT

Predictive maintenance (PdM) has increasingly used data-driven prognostic models to predict equipment degradation and remaining useful life (RUL) in the aerospace industry. In the aerospace industry, predictive maintenance (PdM) has increasingly used data-driven prognostic models to estimate equipment degradation and remaining useful life (RUL). However, most existing methods focus solely on predictive accuracy and offer little support for assessing reliability, explaining, evaluating risk, or making maintenance decisions. The existing drawbacks are overcome by proposing a novel approach to diagnosing and predicting the risk of failure in turbofan engines, based on the NASA C-MAPSS FD001 dataset, within explainable multiagent frameworks for maintenance decision intelligence. The proposed framework is based on a multitask deep learning architecture consisting of a convolutional neural network (CNN), a bidirectional long short-term memory (BiLSTM) network and an attention mechanism to jointly predict the health of the engine and the RUL value. A reliability assessment agent is a fusion of rule-based uncertainty estimation with deep ensemble statistics with a confidence-aware prognostic output. Operational risks are calculated by a Mamdani-type fuzzy risk agent on the basis of the prognostic and degradation indicators, and an explainability agent combines attention analysis, feature importance assessment and feature attribution based on SHAP to provide interpretable evidence. A multiagent reasoning layer also provides information about prognosis to an adaptive multi-criteria decision-making (MCDM) agent, which transforms the prognosis information into maintenance recommendations.

Experimental evaluations on the NASA C-MAPSS FD001 dataset reveal that the proposed framework is effective at providing reliable, interpretable and decision-oriented prognostic intelligence-based predictive maintenance applications, with MAEs of 0.1047 and 0.1036 for health prediction and RUL prediction, respectively; moreover, the RMSE and RMSE are 0.1300 and 0.1287, respectively, and the R^2 scores are 0.7373 and 0.7403 for health prediction and RUL prediction, respectively, with a NASA score of 36.1192.

Keywords—Predictive Maintenance; Remaining Useful Life Prediction; Explainable Artificial Intelligence (XAI); Agentic Artificial Intelligence; Prognostics and Health Management (PHM); Multi-Agent Decision Intelligence; Reliability-Aware Prognostics; Risk-Aware Maintenance Decision Support.

1. Introduction

Given the high complexity of aerospace systems, intelligent prognostics and health management (PHM) solutions are needed to ensure their safety and efficiency in operation. One of the main tasks within the PHM is the prediction of the remaining useful life (RUL) of aerospace systems. Accurate estimation of the degradation and remaining useful life of these systems allows for the maintenance of those components prior to their failure and, therefore, reduces the costs associated with their catastrophic failure. Should any of these components fail within critical industries such as aviation, such failures can have devastating consequences for those industries and their economies.

As such, the aviation industry has begun to rely upon data-driven and intelligent maintenance models to help ensure the safety of flights and passengers. The implementation of sensors in the components of the aircraft has enabled the collection of data related to the health and degradation of those components. Additionally, the use of machine learning (ML) and deep learning (DL) algorithms to analyze the collected data has allowed accurate predictions of the RUL of critical components.

The NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset is among the most extensively used datasets within the field of prognostics and health management (PHM) for turbofan engines. The data collected from turbofan engines under various operating conditions and failure scenarios have enabled researchers to simulate the health of those components and their degradation over time. Additionally, numerous studies have employed this benchmark dataset to propose data-driven prognostic models based upon techniques such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), attention models, ensemble models, and multitask learning (MTL) models to achieve considerable success in monitoring the health of those components and predicting their RUL.

Despite the considerable success that has been reported from the applications of the prognostic models discussed above, there are still limitations to the current implementations of these models. For instance, most of these models are applied to provide an estimate of the RUL of the component to be maintained. However, current maintenance engineers do not typically utilize the RUL and the accuracy of that estimate as the sole information that they use to make a decision regarding the maintenance of that component. Additional information regarding the reliability of that estimated RUL, the risk to the component of that maintenance decision, the severity of the potential degradation of the component if maintenance is not performed, and any other supporting information to that decision is generally required by the engineers before they will perform maintenance on that component. Thus, the accuracy of any estimate of the RUL is inherently limited unless those prognostic models can also make recommendations or provide support for any maintenance-related decisions by those engineers.

Another important challenge is the limited interpretability of many deep learning-based prognostic models. Although deep learning models are able to effectively learn representations of the complex degradation processes of equipment, the reasons behind the decisions of those models are often difficult to understand. This lack of transparency in the decision-making processes of these models can contribute to a lack of trust in the models' recommendations, especially within applications where such trust in the recommendations of the models is critical. Explainable artificial intelligence (XAI) techniques have, therefore, attracted the attention of many researchers within the field of PHM as a means of providing interpretable insights into the predictions of these models. However, many explainability techniques focus solely upon providing explanations for the features used within the models and do not provide explanations for the maintenance intelligence that can be drawn from those model predictions.

In addition to uncertainty and explainability, another important aspect of many prognostic models is their ability to assess their reliability. Maintenance decisions based upon uncertain or unreliable prognostic model predictions can lead to significant risks for the equipment and its operational processes. While there are a number of methods for providing reliability assessments for deep learning prognostic models, the assessments are often provided as additional information outside of the model predictions themselves. Thus, maintenance personnel are still required to manually assess the reliability of model predictions, evaluate the risks that may result from those unreliable predictions, and make maintenance-related decisions according to those assessments.

Recent developments in the field of Agentic Artificial Intelligence (AAI) offer new opportunities for improving the reliability, explainability, and decision-making of prognostic models. Agentic AI systems include multiple AI "agents" that work together to make decisions or take action. Unlike conventional models that provide only prognostic model predictions to end users, agentic AI models can transform those model predictions into intelligence for end users by incorporating steps such as reasoning and evaluating evidence. These capabilities can be beneficial in the area of maintenance in particular, as maintenance decisions have to be made according to a variety of factors regarding the equipment, such as its health, remaining useful life (RUL), the predicted reliability of those estimates, and the risks associated with taking certain maintenance actions.

Motivated by these challenges and opportunities, in this paper, we propose the Risk-Aware Explainable Agentic Prognostic Intelligence Framework (REMAF) for the purpose of performing predictive maintenance for turbofan engines using the NASAC-MAPSS FD001 dataset. The proposed framework incorporates deep learning prognostic model predictions with an agentic intelligence architecture that can support risk and reliability assessments, explainability, degradation reasoning, and maintenance decision support. A multitask deep learning prognostic model uses a convolutional neural network (CNN), a bidirectional long short-term memory (BiLSTM) network, and an attention network to estimate the health and remaining useful life (RUL) of turbofan engines. The outputs of the deep learning model are provided to a collection of intelligent agents within the framework that work together to transform the prognostic model predictions into recommendations for maintenance personnel.

The framework includes a Reliability Agent, a Risk Agent, and a number of Reasoning Agents, such as a Trend Agent, a Stability Agent, a Health Agent, and a Degradation Pattern Agent. Furthermore, an explainability agent utilizes techniques such as attention-based temporal evidence and feature importance to provide explainable outputs from the prognostic system. Finally, a decision agent performs fusion of the prognosticated and reliability data to recommend maintenance actions for the engines. The main contribution of this work is not the creation of a new architecture for predicting RUL but rather the creation of a framework that bridges the gap between prognosis and maintenance decision support. The integration of prognostics, reliability, risk analysis, explainability, and decision support into a unifying framework represents a step forward in the development of intelligent, decision-centric systems for the field of predictive maintenance. Such a framework would be highly valuable in the area of aerospace maintenance.

The primary contributions of this study are threefold. First, a framework utilizing a multitask CNN-BiLSTM-Attention network is developed to estimate the health and RUL of the engine using the NASA C-MAPSS data. Second, a multiagent system is proposed that includes modules for reliability, risk, degradation, explainability, and MCDM. Third, an explainable

and confidence-aware framework is created that integrates attention models, feature importance models, and a human review process to provide trustworthy and decision-centric maintenance support.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the proposed framework. Section 4 presents the experimental results. Section 5 presents a discussion of the results in comparison with related research. Section 6 discusses the limitations of the proposed framework. Finally, Section 7 presents the conclusions of this work, and Section 8 presents suggestions for future research.

2. Literature Review

Over the past few years, Prognostics and Health Management (PHM) has become a fast-emerging field that has moved from conventional condition monitoring to the insightful field of predictive maintenance, which includes the ability to accurately predict equipment deterioration and remaining useful life (RUL). In recent years, the complex degradation patterns of industries, aircraft and other industrial applications in various parts of the world have been studied more frequently using deep learning techniques for modeling. Advanced learning strategies for learning RUL have been proposed, and Gandhudiya and Anil proposed a learning framework that has the potential to capture nonlinear degradation behavior [1]. Similarly, Dong et al. proposed a framework to predict the health of turbofan engines on the basis of maintenance tasks, not simply RUL prediction, that uses the NASA C-MAPSS database, demonstrating the applicability of such maintenance-oriented prognostic intelligence for turbofan engines for prediction other than RUL prediction [2].

Motivated by these properties, a promising prognosis modeling method is deep learning, which can learn the temporal degradation characteristics from multisensor data. To better extract the dependency of the degradation over time, Salini et al. introduced a temporal fusion network with an attention mechanism [3]. In terms of engine health management, scientific machine learning techniques for modeling engine degradation and predicting engine health have advanced by Barry-Straume et al. [4]. The transformer-inspired architecture was quantitatively investigated in sequence modeling for RUL prediction by Shata and Aboelhassan, who clearly reported that the long-range degradation relationships of sequences can be learned by transformer-based architectures [5]. Li and others have added a deep convolutional neural network to the technique of remaining useful for life estimation, in which the feature extraction method using convolution can extract degradation patterns on multisensor measurements to improve the accuracy of prognosis [24]. Learning to perform multiple tasks has also been investigated to simultaneously learn several health indicators and RULs, minimizing the correlations among the various RULs and representations at the various degradation stages [6]. Furthermore, the results of multiple empirically stacked prognostic estimators are more robust for predicting uncertain operating environments [7].

As artificial intelligence-based prognostic systems become increasingly complex, several proposed lines of research have also gained prominence: the rationalization of the systems. With the growing complexity of artificial intelligence-based prognostic systems, several new lines of research have emerged: RATIONALIZATION OF THE SYSTEMS. Lee et al. proposed an anomaly interpretation process for generating understandable explanations for industrial monitoring systems to support maintenance intelligence in a better way for transparency [8]. Recently, several empirical studies on explainable artificial intelligence

(XAI) have been published that seek to increase the levels of user trust, user accountability, and the responsible use of AI in critical safety contexts [10, 19].

Tasks of reliability assessment and estimation of the health index have received a great deal of attention in the field of predictive maintenance research as well. Badawi et al. suggested a comprehensive health index (CHI) method that integrates multiple indicators in their overall health evaluation from a combination of predictive maintenance points of view [11]. This topic has also been studied extensively in the engineering field and can be observed in other mechanical components, including rolling element bearings, harmonic drives and all other critical engineering assets where the importance of confidently balancing maintenance planning can be seen for various engineering applications [12] [13]. Moreover, systems of predictive maintenance based on digital twins and models of residual regime predictive control have predicted the possibility of using intelligent monitoring and optimum maintenance in cyber-physical systems [14] [15].

To increase the robustness of the model in different working scenarios, different methods of feature learning and domain generalization approaches have been studied. To increase the generalization ability of cross-condition RUL prediction under different operating conditions, Wu et al. proposed a graph-aware meta-network method, named RUL2MNETwork [16]. To overcome some of these constraints, Kumar et al. proposed another lightweight Edge-AI system concept for predictive maintenance with explainability and intelligent maintenance recommendations for resource-constrained systems [17]. They are interested in the types and purposes of applications; moreover, there is a growing emphasis on trustworthy and deployment-oriented prognostic intelligence.

In recent years, some work has been carried out on ‘smart decision-making' in predictive maintenance processes. Autonomous robotic agents for operational decision support have been studied for such industrial operations as a prime example of the ability of special-purpose robotic agents to perceive, reason and make adaptive decisions [20]. Pandey et al. suggested an agentic architecture that improves the present AI system by introducing autonomous agents that can think together and collaborate in performing tasks [21]. Similarly, AI-powered foundation and agentic systems capable of intelligent decision-making in various fields were explained by Altintas et al. as an emergent paradigm shift from machine learning to AI-based intelligent decision systems [22], and Shahadat Hossain et al. suggested how AI-based intelligent decision-making systems are evolving from machine learning to foundation and agentic AI in different application fields [23].

While significant progress has been made in functional space prediction accuracy and in explaining these views, several efforts have been dedicated to assessing the reliability of the current research and conducting risk analysis or decision-making of maintenance as functional spaces. All these findings highlight the need for a new integrated prognostic intelligence system, which can estimate both the health status of the engine and the life until the next machine maintenance; in parallel, the reliability associated with the prognosis, the level of risk of operation, the data on engine degradation for interpretation, and coordinated multiagent intelligence, issues actionable/recommendations up to future maintenance based on reliable prognoses.

Table 1 Summary of Existing Prognostic Approaches

Ref.	Methodology	Dataset	Performance Metrics	Key Observations	Our Gap/Novelty
[4]	Multitask CNN + Residual BiLSTM + Attention for joint turbine health and RUL prediction with uncertainty estimation	Real-world engine fleet data	MAE, PICP, MPIW, CWC	Demonstrates effectiveness of attention-based multitask prognostics and uncertainty-aware prediction	Our framework extends prediction outputs into reliability assessment, risk evaluation, explainability, and maintenance decision intelligence
[2]	Condition-aware preprocessing with BiLSTM, XGBoost, LightGBM and health-state classification	NASA C-MAPSS (FD001–FD004)	Accuracy, ROC, Confusion Matrix	Shows maintenance-oriented outputs are more actionable than numerical predictions alone	Our framework jointly predicts Health and RUL while providing risk-aware maintenance recommendations through agent-based reasoning
[3]	Temporal Attention Fusion LSTM (TAF-LSTM) for fault diagnosis and RUL prediction	EV induction motor dataset	Accuracy, RMSE, MSE	Attention mechanisms effectively capture degradation evolution and temporal dependencies	We utilize attention not only for prediction but also for explainability and evidence generation
[17]	Hybrid Edge-Cloud AI with confidence estimation, SHAP explanations, and fault prediction	Industrial machinery dataset	Inference Time, Accuracy, Recall	Demonstrates importance of confidence-aware maintenance systems	Our framework integrates reliability, risk, reasoning, explainability, and MCDM based maintenance decision support in a unified architecture
[8]	Graph-based anomaly detection with correlation-aware explainability	SWaT Industrial Control System Dataset	F1-score, Precision, Recall	Highlights need for interpretable AI systems in critical industrial environments	Our framework provides explainable prognostics coupled with degradation reasoning and

					maintenance recommendations
[21]	Multi-Agent Architecture with autonomous reasoning, planning, and decision support	General Agentic AI Framework	Qualitative Evaluation	Demonstrates effectiveness of coordinated intelligent agents for complex decision-making tasks	We adapt agentic AI concepts specifically for PHM by integrating prognostic outputs, reliability assessment, risk analysis, explainability, and maintenance decision support

Research Gap

The literature reveals several important research gaps, including the lack of maintenance decision support mechanisms, limited integration of explainability with degradation reasoning, inadequate utilization of prediction confidence and reliability information, and the absence of risk-aware prognostic intelligence frameworks. Furthermore, most studies treat prediction, explanation, and decision-making as independent tasks, resulting in fragmented maintenance intelligence, while the application of agentic AI and multiagent coordination in prognosis and health management (PHM) remains largely unexplored. These limitations motivate the development of a unified, explainable, reliability-aware, and decision-centric prognostic framework.

3. Proposed Framework and Methodology

3.1 System Architecture

The Risk-AwareExplainable Agentic Prognostic Intelligence Framework (REMAF) for turbofan engine predictive maintenance using the NASA C-MAPSS FD001 dataset is presented in Fig. 1. The framework extends the capabilities of existing Remaining UsefulLife (RUL) prediction technologies by integrating reliability assessment, risk assessment, explainability, degradation reasoning, and maintenance decision support.

The framework consists of six major layers: the Data Acquisition and Processing Layer, the Multitask Prognostic Layer, the Reliability Assessment Layer, the Risk Assessment Layer, the Reasoning and Explainability Layer, and the Decision Intelligence Layer.

The process begins by acquiring multivariate sensor data from turbofan engines undergoing degradation. The data are converted into a series of fixed-length sequences. The sequences are provided to a CNN-BiLSTM-Attention-based deep learning model that predicts two values: the engine health score and the remaining useful life (RUL) of the engine. The reliability of the engine is assessed through the use of DeepEnsembles, risk is assessed through a series of risk

assessment agents, the degradation of the engine is explained through attention analysis, attention-weighted feature importance, SHAP analyses, and explanation fusion, and maintenance decision support is provided through the adaptive dynamic MCDM decision support model. Thus, the framework aims to provide maintenance intelligence that can be used to support the planning of maintenance activities for turbofan engines.

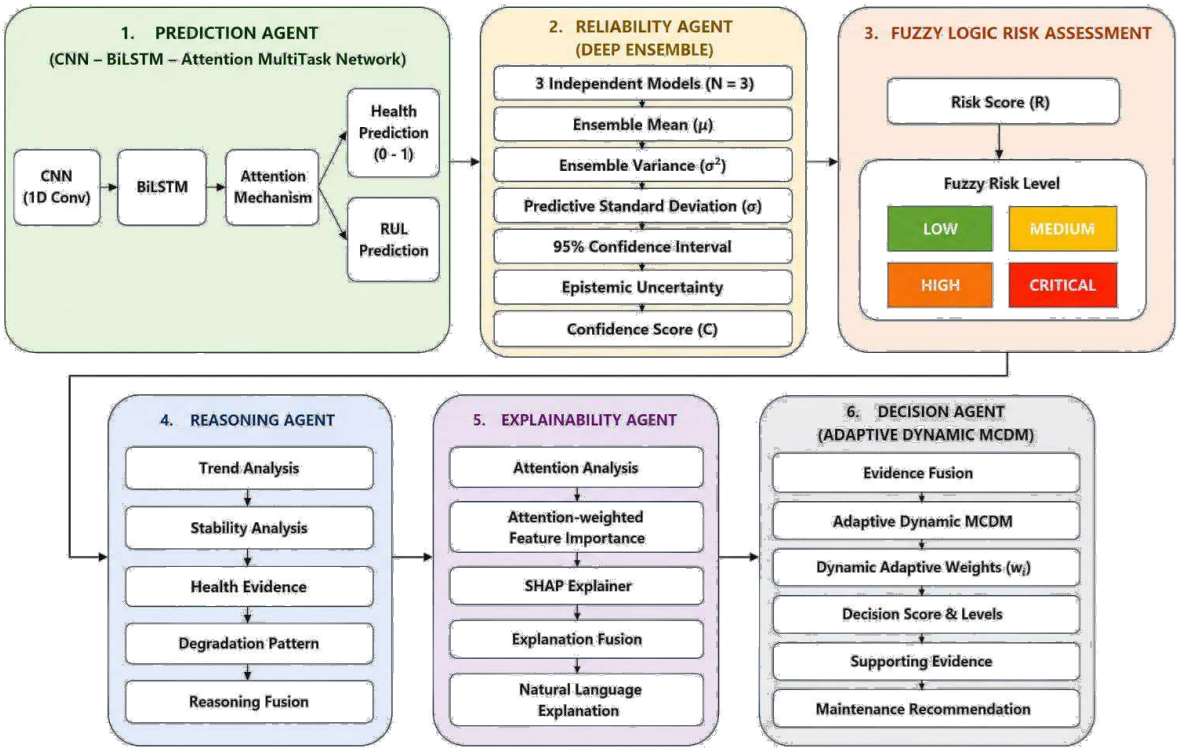

Fig. 1: Overall architecture of the proposed multiagent prognostic intelligence framework.

3.2 Data collection and preprocessing

The proposed framework is based on the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) FD001 dataset. This benchmark dataset provides run-to-failure data for simulated turbofan engines with a single operating condition and single fault mode, where each engine under test is subject to multiple operational cycles that model its degradation from healthy operation to system failure.

The FD001 dataset was chosen because of its controlled degradation nature and its popularity as a benchmarking dataset for prognostic algorithms. Before initial processing, seventeen informative sensor measurements were selected for use in modeling. Inclusive sensor measurements that displayed near-constant behavior (limited sensitivity to degradation) were omitted to increase the robustness of the model and decrease noise. To ensure that the performance of the model is independent of the sensor measurement scale, all the selected sensors were normalized using the min–max normalization method for stable neural network training. Given an input feature x, the normalized value x_norm is given by (1).

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (1)$$

where x_min and x_max denote the minimum and maximum values of the corresponding feature, respectively.

Following the normalization process, temporal sequences were created using a sliding window. A window size of 30 operational cycles was used to represent both short- and long-term trends in the data. Each data sample consists of 30 cycles of 17 sensor measurements each.

The RUL of each engine unit is calculated as the difference between the current cycle and the failure cycle. To enable the deep neural network to learn health-related information, the health score (Healthi) was created using equation (2).

$$\boldsymbol{Health_i} = \frac{\boldsymbol{RUL_i}}{\boldsymbol{RUL_{max}}} \quad \textbf{(2)}$$

RULi refers to the sample at cycle i, and RUL_max refers to the maximum found in the RUL dataset. According to this formulation, healthy engines will have a health value in the vicinity of 1, and very poor-quality engines will have a health value in the vicinity of 0.

3.3 Feature Engineering

Degradation information is important to preserve, while irrelevant variability is removed, a process known as feature engineering. The proposed framework leverages automated feature learning via deep neural networks instead of extensively using manually designed statistical descriptors. The measurements selected for the sensors are arranged into sequences in time that maintain the continuity of degradation. Each input tensor is represented using (3).

$$\boldsymbol{X} \in \boldsymbol{R^{30x17}} \quad \textbf{(3)}$$

where 30 signifies the length of the time segment in the video and 17 signifies the chosen sensor attributes. This representation enables the model to learn the degradation patterns directly from the raw sensor trajectories. The CNN module extracts local degradation signatures, and the interactions between the sensors and the long-term temporal dependency are modeled by the BiLSTM layer. The temporal features are then passed to an attention mechanism to determine the most informative cycles in each sliding window. The context vector generated is a shared vector representation of degradation, which is shared and jointly used by both heads, the health index and the remaining useful life prediction head, allowing multitask learning. The feature engineering process is therefore devoted to maintaining temporal degradation information instead of building handcrafted features, thus automatically learning representations of the degradation that is suitable for the specific task.

3.4 Multitask Prognostic Model Development

The backbone of the REMAF prognostic model is a multitask deep learning model composed of a convolutional neural network (CNN), bidirectional long short-term memory (BiLSTM) and an attention mechanism.

CNN feature extraction layer

The first feature extractor is the CNN. The one-dimensional convolution operation is performed on temporal sensor sequences to detect local degradation signatures and intersensor relationships.

The convolution operation can be expressed using (4),

$$\boldsymbol{h_t} = \boldsymbol{f}(\boldsymbol{W} * \boldsymbol{x_t} + \boldsymbol{b}) \quad \textbf{(4)}$$

Here, W represents the convolution kernel, b represents the bias term and f(.) represents the activation function.

The CNN layer filters out noise and extracts representations that are induced by degradation and are passed to the temporal learning module.

BiLSTM Temporal Modeling Layer

A Bidirectional Long Short-Term Memory network is used to process the extracted feature maps. BiLSTM differs from traditional LSTM, as it can learn temporal dependence in both forward and backward directions, which helps it learn richer degradation representations.

The forward hidden state is calculated and then followed by the backward hidden state, which is concatenated in the form of (5)

$$\boldsymbol{h_t} = [\overrightarrow{\boldsymbol{h_t}}; \overleftarrow{\boldsymbol{h_t}}] \quad \textbf{(5)}$$

This is a bidirectional design that allows the model to use both past and future context information in the sliding window.

Attention Mechanism

While they are effective at modeling temporal dependencies, not all of the operational cycles are equally important for making a prognostic decision. Thus, an attention mechanism is used to detect key degradation durations, thereby enabling the identification of critical degradation durations.

The attention weights are computed using (6)

$$\boldsymbol{a_t} = \frac{\boldsymbol{exp}(\boldsymbol{e_t})}{\sum_{k=1}^{T} \boldsymbol{exp}(\boldsymbol{e_k})} \quad \textbf{(6)}$$

where e_t represents the relevance score associated with cycle t.

The final context vector is obtained using (7)

$$\boldsymbol{C} = \sum_{t=1}^{T} \boldsymbol{a_t h_t} \quad \textbf{(7)}$$

This mechanism allows the model to focus on the most informative degradation stages while dampening out less important temporal information.

Multitask Prediction Heads

The common CNN–BiLSTM–Attention encoder is shared and passed to two task-specific prediction heads, where a common degradation representation is learned prior to being passed to the prediction heads.

The common degradation representation is learned by the shared CNN–BiLSTM–Attention encoder and passed to two prediction heads to perform the respective tasks. The first prediction head is used to estimate the engine health index (HI), which is a measure of degradation conditions, and the second prediction head is used to estimate the remaining useful life (RUL), which is a measure of the number of operating cycles remaining before failure. By doing so, the network can learn degradation representations jointly for both tasks, thereby enhancing the overall prognostic performance. In addition to the health index and remaining useful life predictions, the attention mechanism also produces attention weights that are used to determine the most important operating cycles, which are then used by the explainability agent to interpret the temporal aspects. The second model predicts remaining useful life (RUL), which is the number of cycles to failure.

Combining the optimization of both tasks will lead the network to learn generalized degradation representations that will enhance the overall prognostic performance.

3.5 Reliability Assessment Agent

In safety critical maintenance operations, the reliability of predictions is essential. The proposed deep ensemble-based reliability agent fuses the results of three independent multitask network models to learn the prediction confidence with the ensemble statistics and keep the existing rule-based uncertainty learning as a complementary and adjustable component.

Let the bounded confidence of a prediction x be defined using (8)

$$\boldsymbol{C(x) = 2|x - 0.5|} \tag{8}$$

The uncertainties associated with health and RUL predictions are given by (9) and (10), respectively.

$$\boldsymbol{U_{health} = 1 - C(Health)} \tag{9}$$

$$\boldsymbol{U_{rul} = 1 - C(RUL)} \tag{10}$$

The final confidence score is computed using (11)

$$\boldsymbol{Confidence = 1 - \frac{U_{health} + U_{rul}}{2}} \tag{11}$$

Another use of deep ensemble statistics is to provide a conservative estimate of reliability.

3.6 Risk Assessment and Explainability Layer

Risk Agents

The risk agent is based on a Mamdani-type fuzzy inference system (FIS) that uses multiple degradation indicators to estimate operational risk. The inputs are as follows: Health, Remaining Useful Life (RUL), Confidence, Trend, Stability.

The fuzzy inference engine is used to obtain a continuous risk value, and then this value is mapped to four risk categories as follows: LOW, MEDIUM, HIGH, CRITICAL.

This allows the framework to capture the nonlinearities between degradation indicators and to create understandable operational risk assessments.

SHAP-Based Evidence Generation

The explainability agent has been designed to conform to a multilevel explainability framework that combines attention analysis, attention-weighted feature importance analysis, SHAP-based feature attribution and explanation fusion to offer complete and interpretable evidence to back prognostic predictions.

Attention analysis detects key operation cycles, and the SHAP values are the contributions of each feature to the predictions of health and RUL. The SHAP Explainer provides local and global feature explanations, and the explanation fusion module merges the feature explanations, the attention of time and the feature importance with the SHAP evidence into a holistic explanation in natural language. These explanations are then converted to human readable maintenance evidence.

3.7 Multiagent Reasoning and Decision Intelligence

There are four specialized agents in the reasoning layer:
Trend Agent: This is a service that conveys trends to the user.
Stability Agent: This is a service that provides information on trend stability to the user.
The trend agent is used to determine if the degradation is stable, declining or declining quickly. The Stability Agent checks the consistency of predictions and determines engine behavior as being in high, medium or low stability mode. The health evidence agent leverages explainability outputs from the attention and SHAP analysis modules to identify influential sensor variables, as well as degradation evidence. The degradation pattern agent categorizes the behavior of the operating system as stable operation, degradation or accelerated degradation. All the outputs of the reasoning agents are passed to the adaptive MCDM decision agent.

The final decision score is based on a weighted sum model (Eq. (12)).

$$\boldsymbol{D} = \sum_{i=1}^{n} \boldsymbol{w_i C_i} \qquad \textbf{(12)}$$

Each decision criterion (health risk, RUL risk, confidence risk, trend risk, stability risk, pattern risk, and risk score) is represented by C_i.

The adaptive weights are not fixed weights like those normally used but are variable weights that are changed on the basis of the severity of degradation and the risk conditions. The framework provides one of the following maintenance actions on the basis of the final decision score. On the basis of the final decision score, the framework generates one of the following maintenance actions: NORMAL_OPERATION, MONITOR_CLOSELY, SCHEDULE_MAINTENANCE, IMMEDIATE_MAINTENANCE, HUMAN_REVIEW.

If the level of confidence is below the predefined thresholds, the human review mechanism is triggered to avoid unsafe automated decisions.

3.8 Evaluation Methodology

From various viewpoints, the proposed framework is assessed to achieve a comprehensive assessment. Standard regression measures, such as the root mean square error (RMSE) and mean absolute error (MAE), are used to assess the prognostic performance for both health and RUL prediction tasks.

The reliability assessment framework generates 95% confidence interval (CI) estimates, uncertainty measures and ensemble statistics that are used to evaluate the reliability agent. The actual performance of the risk assessment is measured using the risk scores and risk-level classifications. The explainability performance is analyzed on the basis of feature contribution analysis and evidence generation capability by SHAP. Finally, maintenance recommendation consistency, priority assignment behavior and case based maintenance decision analysis for decision intelligence are measured.

The proposed REMAF framework integrates prognosis prediction, reliability assessment, explainability, risk evaluation, degradation reasoning and adaptive dynamic MCDM based maintenance decision support in a unified framework, which transcends the concept of predictive maintenance from prediction-oriented analytics to intelligent and actionable maintenance decision support.

4. Experimental Results and Discussion

4.1 Experimental setup

The NASA C-MAPSS FD001 turbofan engine degradation data were used to test the proposed Risk-Aware Explainable Agentic Prognostic Intelligence Framework (REMAF). The framework incorporates four agents—two for reliability assessment and two for explainability analysis (via SHAP), an agent for risk assessment, and two multiagent reasoning-based modules—in addition to a CNN-BiLSTM-attention-based multitask prognostic model and a decision agent for adaptive multicriteria decision making (MCDM). The experimental evaluation included not only testing the predictive ability of the framework but also testing the ability of the framework to provide reliable confidence estimates, operational risk analyses, meaningful degradation evidence and actionable maintenance suggestions. The performance was therefore measured in five areas: prognostic prediction, reliability assessment, risk evaluation, explainability and maintenance decision intelligence.

Table 2: Prediction Performance Results

Metric	Health Prediction	RUL Prediction
MAE	0.1047	0.1036
RMSE	0.1300	0.1287
R^2	0.7373	0.7403
NASA Score	-	36.1192

4.2 Prognostic Prediction Performance

The ability of the proposed framework for prediction was first tested for health score and remaining useful life (RUL) prediction tasks. The proposed CNN-BiLSTM-Attention model successfully captures the degradation information from multivariate sensor trajectories, as shown in the obtained results.

The Health Prediction module had a mean absolute error (MAE) of 0.1047, a root mean square error (RMSE) of 0.1300 and an R^2 of 0.7373. Similarly, the RUL prediction module has an MAE of 0.1036, an RMSE of 0.1287, and an R^2 score of 0.7403. The results show that the proposed multitask architecture is able to acquire shared representations of degradation that can be used for both health assessment and lifetime prediction tasks simultaneously.

A score of 36.1192 was achieved for the RUL task with the NASA scoring metric. Given the NASA score, which is more stringent for late predictions than it is for early predictions, the results show the framework's potential for making maintenance-oriented forecast estimates while minimizing the number of potentially hazardous forecasts that are late.

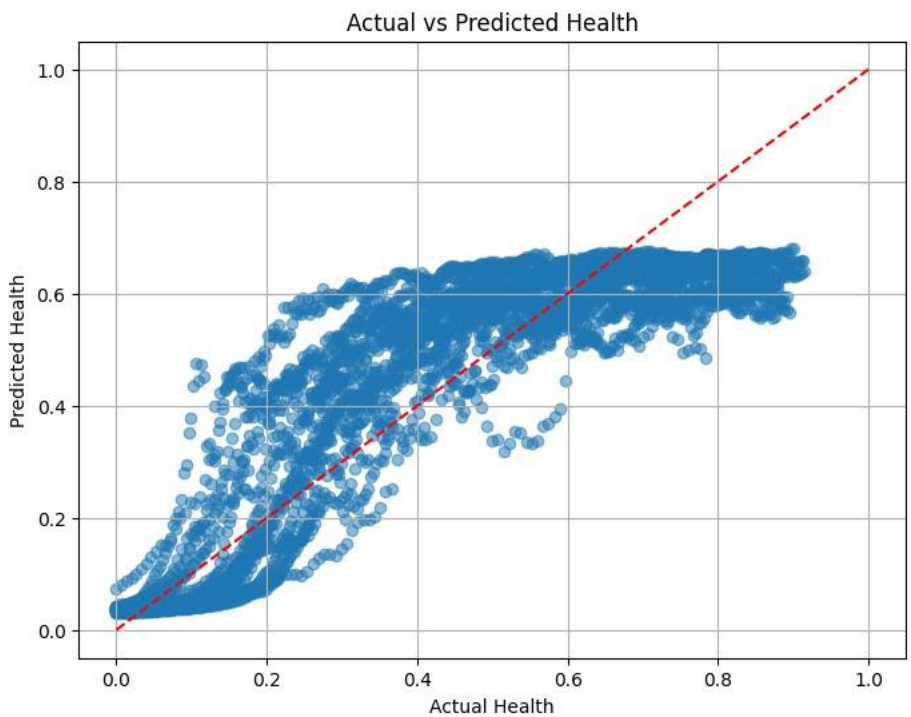

Figure 2. Actual vs Predicted Health

The data in Figure 1 show the relationships between the actual health values and the predicted values. The degradation trajectories are in good agreement between the predictions and measurements, suggesting that this model is suitable for describing degradation evolution throughout the entire engine life cycle. Only slight differences are found under very degraded operating conditions when the variability in the sensors is greater.

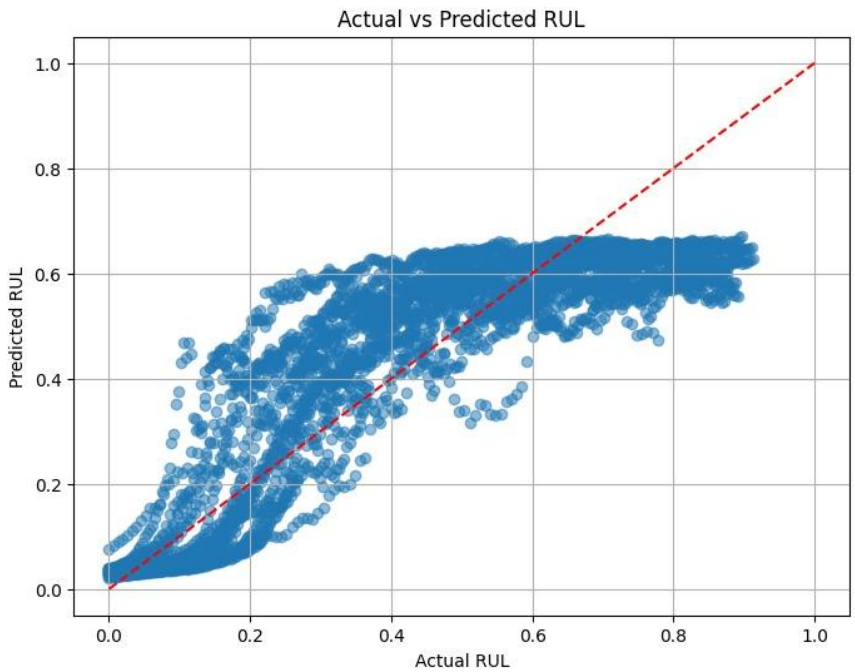

Figure 3. Actual vs Predicted RUL

As shown in Figure 2, the predicted RUL values correlate well with the actual RUL values and are strongly positively correlated. The prediction variance is greater during the intermediate stages of degradation, whereas mild saturation is observed at higher RUL values; thus, the model still strongly agrees with the overall temporal degradation pattern under the different engine operating conditions.

4.3 Reliability Assessment Performance

One of the main goals of the proposed framework is to expand beyond traditional prediction systems by integrating reliability assessment in the maintenance decision-making process. Deep Ensemble Reliability Analysis was incorporated to accomplish this goal.

The average level of confidence produced by the reliability assessment framework on the validation samples was 0.8261, representing a generally high prediction level of confidence produced by the reliability assessment framework. The relatively low uncertainties in prediction (Health: 0.1760; RUL: 0.1717) also demonstrate a relatively low uncertainty in prediction and a consistent level of confidence.

The results show that the reliability assessment framework can offer an extra level of trustworthiness assessment to complement traditional prediction outputs. The framework is not

only able to predict values but can also quantify the level of confidence in the prediction, which can be used in the maintenance decision-making process in a safer manner.

The reliability estimates are also used as important inputs to the risk agent and decision agent, which in turn are used to make confidence-aware maintenance recommendations and triggers for human review when the confidence level is below the acceptable limit.

4.4 Risk Assessment Analysis

The risk agent combines the information of the health score, remaining useful life (RUL), prediction confidence, degradation trend and stability based on a Mamdani-type fuzzy inference mechanism to provide an operational risk estimation.

The framework resulted in an average risk score of 61.45, which indicated that the scores tended to be high overall in the validation set. This is what would be expected since many samples in the validation set would be from engines in the late stages of degradation.

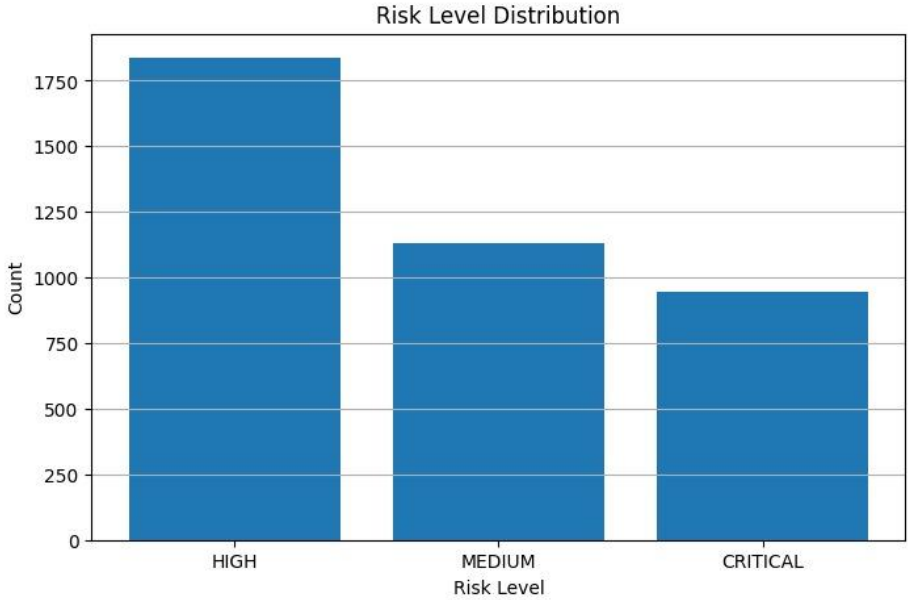

Figure 4. Risk Level Distribution

The distribution of the risk level produced by the risk agent is shown in Figure 3. The proposed fuzzy logic risk assessment agent categorized 1,130 of them as medium risk, 1,837 as high risk and 946 as critical risk. The distribution is a representation of how the dataset, called NASA C-MAPSS FD001, degrades progressively over time and reflects the performance of the agent in finding the engines that need maintenance. If the distribution is analyzed, 46.95% of the samples fall into HIGH risk, and 24.18% fall into CRITICAL risk. Only 28.87% of the observations were given the risk level of MEDIUM. There were no LOW-risk cases among the cases examined, indicating that the validation set is composed mainly of engines in moderate to severe degradation conditions. The results highlight the capacity of the risk agent to effectively recognize the different levels of operational urgency and provide useful hazard categorization for downstream maintenance planning.

4.5 Explainability and Degradation Evidence Analysis

To enhance the transparency and interpretability of the proposed framework, an explainability agent is integrated that leverages attention-based temporal reasoning and feature importance analysis.

Table 3: Explainability Results

Explainability Component	Output
Most Important Cycle	30
Maximum Attention Weight	0.0771
Top SHAP Feature	op_setting_1
Top Positive Contributors	op_setting_1, sensor_4, sensor_11
Most Influential Negative Feature	sensor_15

The explainability agent is a brainchild of the author who uses attention-based temporal reasoning along with feature analysis using SHAP to make the model more explainable in terms of its features and their importance. Cycle 30 was the most influential timestep in terms of attention analysis, reflecting the fact that the operational information in the last few cycles has the greatest impact on the prognostic predictions. The results also revealed that both the operational settings (op_setting_1 and op_setting_2) and the operational measurements (sensor_4, sensor_11 and sensor_15) significantly influence the results of the SHAP analysis, which highlights that the operational parameters are important for both health and RUL prediction. These explanations, combined, provide interpretable evidence to support the maintenance recommendations in the framework.

4.6 Maintenance Decision Intelligence Analysis

The last layer of the framework translates the notions of prognosis, reliability, risk and explainability to actionable maintenance suggestions via the Adaptive MCDM Decision Agent.

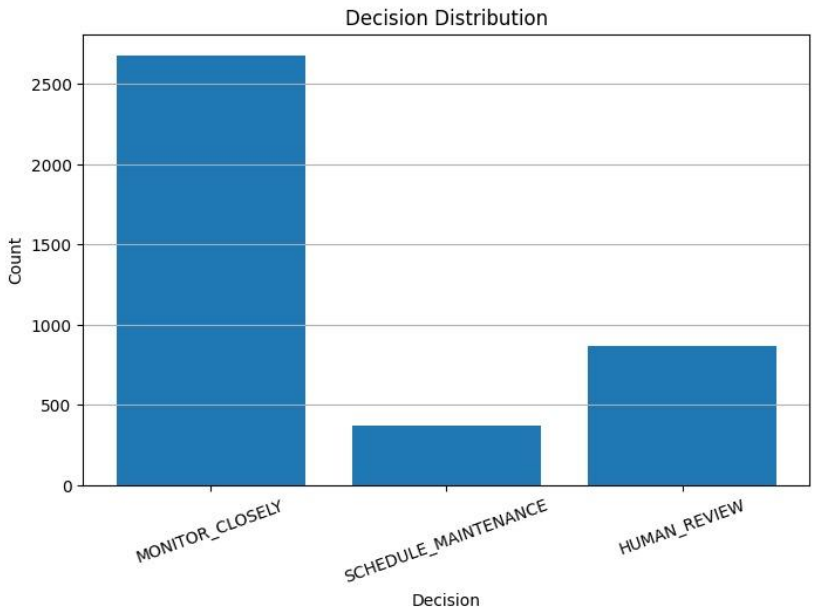

Figure 5. Decision Distribution

For 2,677 engine instances, the Adaptive Dynamic MCDM Decision Agent recommended MONITOR_CLOSELY; 863 engine instances were recommended to be maintained by the human user; 373 engine instances were recommended for SCHEDULE_MAINTENANCE; and no engines were recommended for NORMAL_OPERATION or IMMEDIATE_MAINTENANCE. This distribution suggests that the engines assessed were mostly in a moderate state of degradation and should be continuously monitored or dealt with by an expert rather than intervention or normal use.

The results of the maintenance decisions generated suggest that most samples fall into the category MONITOR_CLOSELY. This behavior implies that there are many engines that have conditions of degradation that need to be monitored more closely without immediate action. There are also a considerable number of samples that have been classified as HUMAN_REVIEW. This illustrates the success of the confidence-aware decision mechanism, with uncertain predictions being escalated to the expert for decision-making and not leading to potentially unsafe automated maintenance actions. Schedule maintenance recommendations are also presented, suggesting that the framework is able to detect the maintenance critical state of engines. None of the samples were deemed to be NORMAL_OPERATION or IMMEDIATE_MAINTENANCE, as the scores in this dataset are representative of samples that were degraded.

4.7 End-to-End Framework Case Study

To illustrate the integrated working of the proposed framework, the application of the framework for the case study of Engine 81 was used. A case study of Engine 81 is presented in this table.

Table 4: Engine 81 Case Study

Component	Output
Health Prediction	0.0378
RUL Prediction	0.0347
Confidence	0.5362
Risk Score	87.83
Risk Level	CRITICAL
Pattern	STABLE_OPERATION
Trend	STABLE
Stability	LOW
Decision	HUMAN_REVIEW

The engine is running very close to failure conditions, as suggested by the Health Prediction (0.0378) and RUL Prediction (0.0347). The corresponding confidence score is much lower than the average of the dataset (0.5362), leading to high uncertainty. Thus, the risk agent gave the engine a risk score of 87.83 and assigned the engine to CRITICAL. The Trend and Pattern Agents reported stable behavior, but the Stability Agent reported stabilities to be at a low level, indicating that the observed state might not be a good indicator for making a fully automated maintenance decision. Therefore, the decision agent developed a HUMAN_REVIEW recommendation. These results demonstrate the necessity of considering both prediction confidence and risk evaluation and reasoning information when a decision is being made.

The framework not only takes the degradation scale into account but also takes uncertainty into account before providing maintenance recommendations.

4.8 Discussion

The experimental results presented in this paper demonstrate the extension of conventional prognostic systems beyond the scope of prediction. The CNN-BiLSTM-Attention architecture effectively delivers HealthIndex and RUL predictions. Furthermore, the deep ensemble-based reliability agent contributes reliability intelligence to the system by providing confidence in the predictions made by the CNN-BiLSTM-Attention architecture.

The fuzzy logic risk agent translates the predictions of the main model into risk assessments for the machine components. The reasoning agent determines degradation trends in the components of the engine. The explainability agent provides interpretability of the main model by providing attention analysis, attention-weighted feature importance, SHAP analysis, and explanation fusion. Finally, the adaptive dynamic MCDM decision agent integrates the intelligence of all the other agents and provides intelligent maintenance decision-making recommendations for the engines. An Engine 81 case study demonstrates the integration of all these components and their contribution to intelligent maintenance decision-making. These results demonstrate that the proposed framework is not just a prognostic model but also an agentic predictive maintenance intelligence system that can perform intelligent maintenance decision-making for machinery operations.

5. Comparison and Validation

A representative set of studies from the areas of prognostics, predictive maintenance, explainable AI and agentic intelligence was used for comparison with the proposed Risk-Aware Explainable Agentic Prognostic Intelligence Framework (REMAF). Previous studies have shown very good skills in specific modules, including the prediction of remaining useful life (RUL), classification of health states, estimation of uncertainty, explainability, and autonomous decision-making. Most of the approaches, however, are used for individual components, and they do not offer a single framework of maintenance intelligence. The focus of scientific machine learning (SML) engine health management approaches is mostly on health and risk-aware RUL prediction with uncertainty estimation without considering risk-aware maintenance reasoning or decision support. Similarly, the multiclass predictive maintenance framework for Jet Engines does not consider confidence-aware decision-making or reliability assessment for maintenance. Prognostic models based on attention can capture the temporal degradation pattern well, but the outputs of the models do not generally include maintenance intelligence integration and end at prediction generation.

Although explainable predictiveness maintenance systems and edge-AI frameworks have been recently developed, they do not address operational risk assessment, degradation reasoning or adaptive maintenance recommendation generation at the feature level or provide confidence estimation. Similarly, anomaly interpretation frameworks offer clear explanations but are limited to identifying the fault and not planning for prognosis. While not explicit in their design to perform Prognostics and Health Management, agentic AI architectures illustrate a high level of reasoning and autonomous decision-making. In contrast, the proposed REMAF is a single architecture that combines health prediction, RUL prediction, the reliability assessment agent, risk quantification, SHAP-based explainability, multiagent degradation reasoning, and adaptive MCDM decision support. The experimental results show good competitive prognostic performance (RUL RMSE = 0.1287, $R^2 = 0.7403$), and they are able to provide confidence-based, risk-based and interpretable maintenance recommendations. This capability is both common and unique to the proposed framework compared with existing prediction-centric approaches and demonstrates the effectiveness of the proposed framework as a comprehensive predictive maintenance intelligence system.

6. Limitations

The proposed framework was tested only on the NASA C-MAPSS FD001 dataset, and the generalization ability of the framework across multiple operating conditions is unproven. While the deep ensemble reliability agent and SHAP-based explainability agent enhance prediction reliability and transparency, they do not account for all sources of uncertainty or offer a full root-cause analysis. The fuzzy logic risk agent and adaptive dynamic MCDM decision agent, however, have yet to be tested in actual industrial maintenance environments and rely on fixed rules.

7. Conclusion

This paper presents a risk-aware explainable agentic prognostic intelligence framework (REMAF) for turbofan engine predictive maintenance using the NASA C-MAPSS FD001 dataset. Conventional Prognostics and Health Management (PHM) systems focus on the prediction of the remaining useful life (RUL) of the engines. However, they do not offer assessments of the reliability of the components, the level of risk that the components pose to the turbofan engines, explanations of the maintenance decisions, or the intelligence behind these maintenance decisions. Thus, despite the availability of PHM systems, their lack of these features limits their applicability in the field. The proposed framework includes a CNN–BiLSTM–Attention-based MultiTask model to simultaneously predict the health index (HI) and the remaining useful life (RUL) of the turbofan engines. A deep ensemble model is used to create a reliability agent to provide information on the confidence of each prediction. Furthermore, an explainability agent utilizes attention analysis, attention-weighted feature importance, the SHAP explainer, and explanation fusion to provide explanations for each predicted value for the health index and remaining useful life. A fuzzy logic risk assessment model provides information on the risks that the engines pose to the turbofan engines. Additionally, a multiagent degradation reasoning model explains the degradation of the turbofan engines, and an adaptive dynamic MCDM decision agent creates intelligent maintenance recommendations using the information from the other agents in the framework. The experimental results using the NASA C-MAPSS FD001 dataset show that the framework is effective. For the MultiTask model, the mean absolute error (MAE) values for the health index and remaining useful life were 0.1047 and 0.1036, respectively, and the coefficient of determination (R^2) scores were 0.7373 and 0.7403, respectively. The deep ensemble model achieved an average confidence score of 0.8261, and the fuzzy logit risk assessment model produced risk assessments with an average risk score of 61.45. Furthermore, the explainability agent found influential operating cycles and features for degradation and generated explanations for each turbofan engine using the included methods. Finally, the adaptive dynamic MCDM decision agent created maintenance decisions for each turbofan engine using the information from the other agents in the framework. Thus, the proposed framework successfully transforms the focus of turbofan engine maintenance from the prediction of health and remaining useful life to a transparent, explainable, and intelligent maintenance paradigm. By including these various AI models within a single framework, this study highlights the potential use of AI in the aerospace industry for PHM systems.

8. Future Work

Despite the promising results obtained with the proposed framework, several opportunities for enhancing the system further exist. One of the crucial aspects that can be incorporated into the system in the future is the integration of degradation propagation models based on the physics of the engine. Such an approach can improve the robustness of degradation predictions under scenarios with unseen engine conditions or limited data availability. Furthermore, it may improve the interpretability of the model's suggestions for maintenance. Another important aspect of future research would be the investigation of techniques to perform root-cause analysis of the degradation of the components of the engine. Although the explainability tools currently implemented can identify which features of the engine are the most critical in determining degradation, there is the potential for creating techniques that determine which component degradation results in abnormal sensor readings from the engine.

Furthermore, it may also be interesting to investigate the causes of these failures. A deeper understanding of the cause of engine failure will allow maintenance agents to better understand these failures and incorporate this understanding into their suggested maintenance procedures. Furthermore, another interesting direction for future research would be to incorporate some form of feedback learning between the various intelligent agents within the framework. The existing framework utilizes a unidirectional information flow between intelligent agents. However, incorporating feedback capabilities into agents such as the risk agent, the explainability agent, and the decision agent will allow their future decisions to be informed by the previous decisions of the other agents. This form of feedback may further improve the system's intelligence.Finally, the current experiments were performed on the NASA C-MAPSS FD001 data subset only. It would be interesting to extend the experiments to other data subsets, specifically FD002-FD004 and other datasets that represent other industrial engines. These suggested extensions to the framework have the potential to further develop the system into an even more intelligent, explainable, and deployable next-generation predictive maintenance intelligence system.

9. Declarations

Ethical Approval

Not applicable.

Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Funding

The authors received no specific funding for this research from any public, commercial, or not-for-profit funding agency.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

Naman Bafna: Conceptualization, Methodology, Software, Investigation, Data Curation, Writing – Original Draft.
Pranav Kumar D: Validation, Experimental Analysis, Visualization, Writing – Review & Editing.
Prathap Rudra Boppuru: Supervision, Conceptualization, Project Administration, Writing – Review & Editing.

Data Availability

The data supporting the findings of this study are publicly available through the **NASA C-MAPSS turbofan engine degradation dataset** hosted by the NASA Prognostics Center of Excellence.

Dataset:
https://data.nasa.gov/dataset/c-mapss-aircraft-engine-simulator-data

Any additional processed data generated during the study are available from the corresponding author upon reasonable request.

Code Availability

The implementation code developed for this study is available from the corresponding author upon reasonable request.

Acknowledgements

The authors gratefully acknowledge the **NASA Prognostics Center of Excellence** for providing the publicly available C-MAPSS benchmark dataset used in this research.

AI Statement (Recommended for Springer)

The authors declare that generative artificial intelligence (AI) tools were used only to assist with language refinement and manuscript editing. The authors take full responsibility for the originality, accuracy, interpretation, and integrity of all scientific content presented in this manuscript.

REFERENCES

[1.] Gandhudi, M., & Anil, G. R. (2026). Quantum Annealing Enhanced Reinforcement Learning for Accurate Remaining Useful Lifetime Prediction. arXiv preprint arXiv:2606.18503.https://doi.org/10.48550/arXiv.2606.18503

[2.] Dong, B., Zhang, X., Hou, L., Yan, C., Feng, Y., Zhu, W., & Lin, L. (2026). A Multi-Class Predictive Maintenance Framework for Jet Engines Using the C-MAPSS Dataset. Machines, 14(6), 695.https://doi.org/10.3390/machines14060695

[3.] Salini, S. M., Geetha, S., Kabilan, K., & Sumesh, P. (2026). Attention-enhanced deep temporal fusion network for cloud-IIoT prognostics of EV induction motors under dynamic loads. Scientific Reports. https://www.nature.com/articles/s41598-026-51447-5

[4.] Barry-Straume, J., Son, C., Sandu, A., Burke, G., Sundararajan, R., Rimell, A., & Steinrock, J. G. (2026). Scientific Machine Learning for Engine Health Management and Remaining Useful Life Prediction. arXiv preprint arXiv:2605.30593. https://doi.org/10.48550/arXiv.2605.30593

[5.] Shata, S., & Aboelhassan, A. (2026). Toward Decision Transformer for Remaining Useful Life Prediction: A Sequence Modeling Based Approach. IEEE Access. https://doi.org/10.1109/ACCESS.2026.3704028

[6.] Belaunzaran, X., Nappa, A., Artetxe, A., & Sierra, B. (2026). Bifurcated Remaining Useful Life Prediction: A Hybrid Approach for Realistic Uncertainty Characterization. https://doi.org/10.48550/arXiv.2605.31241

[7.] Karlsson, J., Karlsson, A., Bandaru, S., & Bekar, E. T. (2026). Empirical Likelihood Stacking for Remaining Useful Life Estimation. In International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (pp. 303-317). Springer, Cham.https://doi.org/10.1007/978-3-032-28994-0_22

[8.] Lee, S., Lim, S., & Lee, T. (2026). Automated Identification and Interpretation of Anomalous Cases in Industrial Control Systems. Electronics, 15(12), 2705. https://doi.org/10.3390/electronics15122705

[9.] Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical systems and signal processing, 104, 799-834. https://doi.org/10.1016/j.ymssp.2017.11.016.

[10.] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778 .

[11.] Badawi, M., Ibrahim, S. A., Mansour, D. E. A., El-Faraskoury, A. A., Ward, S. A., Mahmoud, K., ... & Darwish, M. M. (2022). Reliable estimation for health index of transformer oil based on novel combined predictive maintenance techniques. IEEE access, 10, 25954-25972. https://doi.org/10.1109/ACCESS.2022.3156102

[12.] Li, Z., Mantas, P., Verwimp, T., Ojo, D. S., Goossens, S., Mauricio, A., ... & Gryllias, K. Fault size and RUL estimation of rolling element bearings using FBG-based strain measurements. Available at SSRN 6927754. https://dx.doi.org/10.2139/ssrn.6927754

[13.] Guida, R., Bertolino, A.C., De Martin, A. et al. Preliminary PHM System Development for Harmonic Drives in Electromechanical Flight Control Systems. Aerotec. Missili Spaz. (2026). https://doi.org/10.1007/s42496-026-00323-6

[14.] Li, H. (2026). Digital Twin-Driven Predictive Maintenance Framework for Complex Mechanical Systems in Industry 4.0. European Journal of Engineering and Technologies, 2(2),41-55.

[15.] Vasilev, A. E. I. Residual-Regime Markov Modeling for Predictive Control in Cyber-Physical Systems. https://doi.org/10.47191/ijcsrr/V9-i6-28

[16.] Wu, Z., Chen, Q., Wei, Z., Ling, Y., Hou, G., Zhao, P., & Zhou, Y. (2026). A graph-aware meta network for domain generalization in cross-condition bearing RUL prediction. International Journal of Metrology and Quality Engineering, 17, 11. https://doi.org/10.1051/ijmqe/2026009

[17.] Kumar, G., Terauds, M., Ajayakumar Raji, A., Semenako, J., Smolaninovs, V., Sics, P. E., & Malayidinja Poikayil Thankappan, A. K. (2026). Design of a Lightweight Edge-AI System for Predictive Maintenance on ESP32-S3. Applied Sciences, 16(11), 5287. https://doi.org/10.3390/app16115287

[18.] Xu, Weiyan and Ma, Qinglong and Tu, Jielei and Xu, Jiyong, Empirical Verification and Performance Analysis of Divide-and-Conquer Learning for Highly Mutant Agricultural Load Forecasting. https://dx.doi.org/10.2139/ssrn.6960649

[19.] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012 .

[20.] Al-Zadjali, I. (2026). The application of autonomous AI agents in upstream oil and gas supply chain decision-making. Volume-10 Issue 06, June -2026. ISSN: 2456-9348. Impact Factor: 8.232.

[21.] Pandey, V., Jaiswal, N., Ansari, M. H., & Siddiqui, F. Agentic AI Systems: Architecture, Applications, and a Case Study of OpenClaw. https://doi.org/10.63920/tjths.52016

[22.] Altintas, M., Duru, I., Yazici, I., Karahan, S. N., & Cimen, S. (2026). Beyond Connectivity via Artificial Intelligence Agents Enabling Perceptive, Adaptive, and Anticipatory 6G Networks. IEEE Access, 14, 45977-46023. https://doi.org/10.1109/ACCESS.2026.3667803

[23.] Shahadat Hossain, Md Risalat Hossain Ontor, & Md Golam Sarwar. (2026). From Machine Learning to Foundation and Agentic AI: Evolution of Intelligent Decision Systems Across Domains. Journal of Computer Science and Technology Studies, 8(7), 112-126. https://doi.org/10.32996/jcsts.2026.8.7.9

[24.] Li, X., Ding, Q., & Sun, J. Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. Reliability Engineering & System Safety, 172, 1-11. https://doi.org/10.1016/j.ress.2017.11.021 .