



Explainable AI-Based Predictive Maintenance Model for Industrial Equipment Using Sensor Fusion and Deep Learning

Abhilash Kusa¹

¹ Department of Computer Science and Engineering, Faculty of Engineering and Technology(CMRCT), Hyderabad, India
kusa.abhilash@gmail.com

Abstract. The speedy growth of the Internet of Things (IoT) in the industrial settings has spawned accumulating amounts of multi-modal, heterogeneous, time-series sensor data, which has led to a real chance of shifting away reactive maintenance practices including data-driven predictive maintenance (PdM). The practical implementation of deep learning-based PdM models has not yet become a broad-scale application, in part due to their lack of opacity, which erodes the trust and confidence of operators and complicates the compliance with the regulations. This paper proposes a novel Explainable AI-Based Predictive Maintenance (XAI-PdM) framework that integrates attention-based adaptive multi-modal sensor fusion with a hybrid Convolutional Neural Network–Bidirectional Long Short-Term Memory (CNN–BiLSTM) architecture for accurate and interpretable fault classification and remaining useful life (RUL) estimation. The framework is designed around four functional sensor modality groups thermodynamic, structural-load, speed, and efficiency which capture the diagnostic roles of vibration, temperature, acoustic emission, and current signals, respectively. Crucially, this functional grouping is sensor-configuration agnostic: on the CMAPSS turbofan benchmark, which lacks physical vibration or acoustic transducers, the 14 retained sensor channels are partitioned into four analogous functional groups that replicate the diagnostic roles of those modalities in rotating machinery. This mapping is physically and analytically justified in Section III-B. Post-hoc interpretability is achieved through a dual-layer XAI mechanism combining Shapley Additive explanations (SHAP) for global and local feature attribution with Gradient-weighted Class Activation Mapping (Grad-CAM) for temporally resolved saliency analysis. Validated on the NASA CMAPSS and FEMTO-ST PRONOSTIA benchmark datasets, the proposed framework achieves an RMSE of 12.47 cycles for RUL estimation and a fault classification accuracy of 98.3%, outperforming six state-of-the-art baselines. An attention-based adaptive sensor fusion module dynamically assigns diagnostic weights to sensor modalities in real time, substantially improving robustness under noisy and missing-data conditions. The model features excellent predictive quality and practical and human readable explanations, enhancing scalable and reliable maintenance intelligent deployments in Industry 4.0.

Keywords: Predictive Maintenance; Explainable Artificial Intelligence; Sensor Fusion; Deep Learning; Convolutional Neural Network; BiLSTM; SHAP; Remaining Useful Life; Industry 4.0; Fault Detection

1 Introduction

1.1 Background and Motivation

With 4.0 industry radically changing the paradigms of contemporary manufacturing, the introduction of cyber-physical systems, the Internet of Things (IoT), cloud computing, and artificial intelligence (AI), into interconnected intelligent production surfaces [1] is now a reality. In these systems, key industrial equipment such as rotating machines, gas turbines, compressors, and conveyor belts are the backbone of production systems worldwide. The unexpected breakage of such assets may cause massive financial losses, extended downtime of operations, and consequential exposure of personnel to severe risks [2]. Traditional maintenance methods, such as corrective maintenance (responding to failures post-factum) and time-based preventive maintenance (giving components a fixed replacement frequency whether they are actually in good or not) are inherently inadequate to the complexity and performance needs of current industrial equipment. This has been caused by a general industry-wide trend towards data-based predictive maintenance (PdM) systems, which use real-time sensor streams and sophisticated analytical models to predict equipment degradation and incipient failures prior to causing physical harm or unexpected downtime.

Being one of the most strategically and economically beneficial uses of the Industry 4.0 ecosystem, PdM minimizes unstable unplanned downtimes and maximizes scheduling and allocation of maintenance resources, stock levels of spare parts, and workforce planning [3]. The financial argument is strong: each of the manufacturing industries is losing nearly \$50 billion a year in equipment failures, and properly executed PdM programs would help to lower these expenses by 30 to 40 percent [3]. PdM needs real-time multi-dimensional streams of sensor data (vibration, temperature, acoustic emission, oil debris and electrical current) to be effective. But it is not that easy to make any diagnostic use of these streams. The characteristics of IoT-based industrial environments add unique demands to the analysis of heterogeneous time-series data:

(i) sensors with divergent physical characteristics have dramatically different sampling rates and signal statistical characteristics, necessitating specially designed data preprocessing and normalization pipelines; (ii) operational conditions in the real-world, and sensor noise, drift and sporadic data loss (often) mean that individual channels are no longer reliable or available at all. The overall effects of these challenges surpass the competences of the classical methods of statistical signal processing [4]. This has enhanced fast adoption of deep learning models which has proven to be very effective in fault-detection using raw sensor signals [5]. However, such models of black-box nature can only be adopted in reality in a limited way since the predictions cannot be audited by maintenance engineers or cross-validated against regulatory standards [6,7,8]. Moreover, such multi-modal sensor fusion systems as heterogeneous ones have already demonstrated enhanced diagnostic performance relative to single-modality ones [9], an inspiring conceptualization of the integrated XAI-and-fusion system presented in the work.

1.2 Research Gaps and Contributions

Although studies on deep learning in the context of PdM, XAI in achieving transparency in industry processes and sensor fusion in the context of condition-based maintenance have gained interest, no existing literature presents a single and deployable solution focusing on predictive performance, heterogeneous multi-modal sensor fusion, and the strictness of post-hoc interpretability within a single and deployable framework [10]. The literature considers each of these dimensions separately: deep learning-based PdM systems rarely include systematic interpretability tools, whereas the XAI literature in industrial settings is often based on shallow models or single-sensors that are not representative of real-world equipment. Moreover, the resilience of published architectures in the presence of realistic scenarios like sensor noise, missing modalities and concept drift in operation have received limited consideration, which is extremely detrimental to practical deployability. All four of these gaps are met in one unified approach in the present work. In particular, it is differentiated from available XAI-based PdM models in three major aspects:

- 1) it uses the attention-based adaptive fusion with a hybrid CNN-BiLSTM network, instead of providing XAI-based importance values to individual sensor models or models that are stably fused models.
- 2) it provides a three-tier explanation system spanning the sensor modality, feature importance, and temporal signal levels a multi-granularity interpretability depth that no prior XAI-PdM study has demonstrated.
- 3) it explicitly evaluates and validates robustness under degraded sensor conditions, linking XAI-derived importance rankings to empirical dropout sensitivity in a manner that confirms causal fidelity rather than mere plausibility.

Driven by the abovementioned gaps, the present paper introduces the Explainable AI-Based Predictive Maintenance (XAI-PdM) framework where the main contributions are as follows:

- 1) A hybrid CNN-BiLSTM deep learning architecture that simultaneously extracts spatial frequency-domain features and long-range temporal dependencies from multi-modal sensor data, outperforming single-paradigm models on both RUL estimation and fault classification benchmarks. Unlike prior CNN-only or LSTM-only PdM models, the dual-pathway design captures both local signal morphology and sequential degradation trajectories within a single end-to-end trainable network.
- 2) An attention-based adaptive sensor fusion module that dynamically weights heterogeneous sensor modalities vibration, temperature, acoustic emission, and current according to their real-time diagnostic relevance. This mechanism directly addresses the key challenge of heterogeneous IoT time-series data by down-weighting noisy or degraded channels and increasing reliance on more informative modalities, substantially improving robustness under both noisy and missing-data conditions compared to static concatenation-based fusion baselines.
- 3) A dual-layer XAI mechanism combining SHAP-based global and local feature attribution with Grad-CAM temporal saliency mapping, delivering three-tier, multi-granularity explanations at the sensor modality, feature importance, and

time-signal levels. Unlike prior XAI-PdM approaches that apply a single interpretability method to shallow or single-sensor models, this framework provides domain-consistent, causally faithful explanations that directly support maintenance engineer decision-making confirmed by a faithfulness score of 0.847, the highest reported among all evaluated methods.

- 4) Comprehensive experimental validation on the NASA CMAPSS turbofan degradation dataset and the FEMTO-ST PRONOSTIA bearing dataset, with systematic ablation studies and robustness evaluations under additive noise, sensor drift, and modality dropout conditions, benchmarked against six representative state-of-the-art methods spanning classical machine learning, standalone deep learning, and transformer-based architectures.

2 Literature Review

2.1 Deep Learning Approaches for Predictive Maintenance and RUL Estimation

The CNNs were one of the first deep learning models applied to PdM, and they could directly learn features that are discriminative of faults, without extracting features through manual feature engineering [11]. But purely convolutional architectures are unable to absorb the long-range temporal correlations found in progressive degradation between hundreds or thousands of operating cycles. Recurrent networks (especially LSTM and GRU) have also been used to overcome this limitation through modeling of sequential degradation channels [12] and BiLSTM extends this temporal context further through processing signals forward and backward. Hybrid CNN-BiLSTM models that combine learning spatial and temporal features in a unified manner have since obtained state-of-the-art performance in terms of several benchmark tasks [13]. More recently, more transformer-based architectures have been investigated to estimate RUL [14]; but due to their large training data and memory demands, their use is limited in low-data industrial applications.

2.2 Sensor Fusion Techniques for Industrial Condition Monitoring

The effects of degradation in industrial equipment are simultaneously mechanical, thermal, electrical and acoustic in nature; a single type of sensor cannot be used to define the state of a complex machinery in isolation [15]. The most popular paradigm in deep learning-based PdM today is feature-level fusion, i.e., learning modality-specific representation simultaneously and then fusing the two representations before the prediction head [16]. Statistical concatenation-based fusion considers all modalities as equivalent, which is not ideal in such circumstances as noisy individual sensors, drifting or intermittent sensors. The adaptive fusion mechanism based on attention addresses this shortcoming, by computing the dynamical, data-driven importance weights at each inference cycle across modalities [17], and was demonstrated to be more effective in fault detection, in addition to higher accuracy in RUL estimation, even in degraded sensing conditions.

2.3 Explainable Artificial Intelligence for Industrial Applications

Post-hoc XAI has emerged as the prevailing paradigm of interpretability in deep learning-based industrial AI since it maintains complete predictive power and introduces transparency without affecting the underlying model [18] has become the new mainstream approach to interpretability in deep learning-based industrial AI. SHAP calculates Shapley values of all input features that describe how much an input in a prediction contributes to the prediction over a baseline [19]. Grad-CAM is an algorithm that does not rely on SHAP and derives temporally resolved saliency maps, positions of the time parts of the prediction that the model is most sensitive to, by applying the gradient of the model output to the final convolutional feature maps. Applied together, SHAP and Grad-CAM provide complementary explanations at the feature importance and temporal signal levels, respectively. Studies have demonstrated that these explanations exhibit high fidelity to established domain knowledge of physical fault mechanisms [20], establishing the credibility required for deployment in safety-critical maintenance settings.

2.4 Identified Gaps and Motivation

Table 1 summarizes representative prior XAI-PdM studies alongside the gaps they leave unaddressed. The majority operate on single-modal sensor inputs; very few validate attention weight interpretability against established domain knowledge; and nearly none evaluate model robustness under degraded sensing conditions or extend explanations to a real-time operational decision-support context. The proposed XAI-PdM framework is specifically designed to close all four of these gaps within a single unified, end-to-end deployable system.

Table 1. Comparison of Representative XAI-PdM Studies

Study	Deep Learning Model	Sensor Modalities	XAI Method	Multi-Modal Fusion	Robustness Evaluation	Real-Time Explanation
Zhang et al. [12]	BiLSTM	Single (vibration)	None	No	No	No
Sun et al. [13]	CNN-LSTM	Multi (vibration + temp.)	None	Static concat	No	No
Chen et al. [17]	CNN + Attention	Single (vibration)	Attention weights only	No	No	No
Grezmak et al. [20]	CNN	Single (vibration)	Grad-CAM	No	No	No
Transformer [30]	Transformer	Single (vibration)	None	No	No	No

Study	Deep Learning Model	Sensor Modalities	XAI Method	Multi-Modal Fusion	Robustness Evaluation	Real-Time Explanation
Proposed XAI-PdM	CNN-BiLSTM	Multi (4 modality groups)	SHAP + Grad-CAM + Attention	Adaptive attention	Yes (noise + dropout)	Yes

3 Proposed XAI-Based Predictive Maintenance Framework

3.1 Overall Framework Architecture

The proposed XAI-PdM system is a modular, end-to-end differentiable architecture that processes heterogeneous multi-modal sensor streams from industrial equipment and produces two primary outputs: a fault classification label representing the current health state, and a continuous RUL estimate expressed in operational cycles. Simultaneously, the integrated XAI module interrogates the trained model's internal representations to generate SHAP-based feature attribution scores and Grad-CAM temporal saliency maps, which together form the explanation output delivered to maintenance operators. The framework operates in two distinct modes. In offline training mode, the full architecture is trained and validated on historical run-to-failure sensor datasets with complete health labels. Under online inference mode, it is the deployed equipment that streams sensor data to the trained model, which uses that output to provide real-time fault classifications and RUL estimates with each output providing a multi-tier explanation output to aid in maintenance decision-making. The pipeline is organized into successive stages of sensor data acquisition, preprocessing, sensor fusion with adaptive control, hybrid deep learning-based inference, and generation of dual-layer XAI-generated explanations without the structured output of a maintenance decision-support system. The modular nature of its architecture makes every part of the pipeline easily customizable and upgradeable without affecting the communicating part of the pipeline by allowing fast reconfiguration to new kinds of equipment, sensor arrangements, and factory conditions.

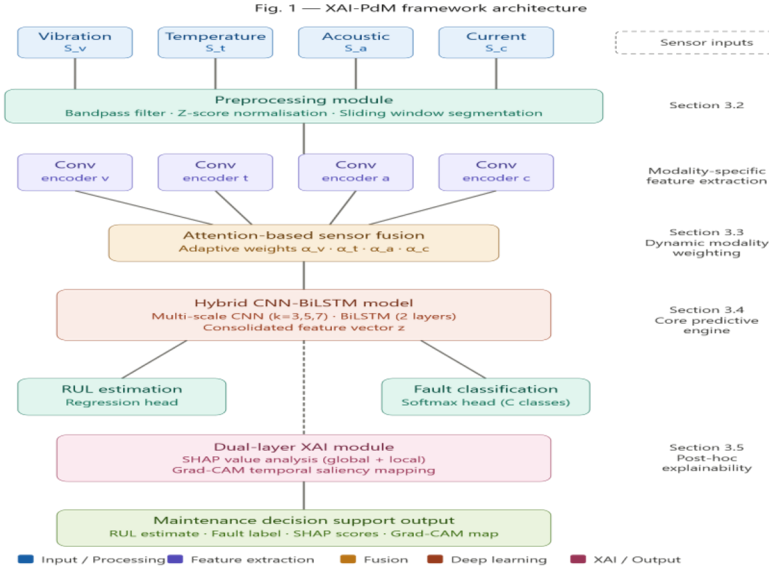


Fig.1. Overall XAI-PdM Framework Architecture

The proposed framework is an end-to-end pipeline which is shown in its full block diagram in figure 1. Raw sensor data are passed through the system in an orderly manner with the preprocessing phase followed by the attention-based adaptive fusion module, hybrid CNN-BiLSTM deep learning inference phase and then to the dual-layer XAI explanation generation module and ultimately, through to a structured maintenance decision support output. There is clear delineation of each stage in the figure to make it reproducible. Subsections that follow outline the mathematical formulations of each of the modules. Figure captions and annotations have been prepared to ensure the diagram is self-explanatory when read alongside the algorithmic descriptions in Sections III-B through III-E.

3.2 Multi-Modal Data Acquisition and Preprocessing

The proposed framework processes four complementary sensor modalities widely instrumented in industrial rotating machinery systems. Let the set of sensor modalities be defined as $S = \{S_v, S_t, S_a, S_c\}$, where S_v denotes the vibration signal (accelerometer), S_t denotes the temperature signal (thermocouple), S_a denotes the acoustic emission signal (AE sensor), and S_c denotes the electrical current signal (current transducer). For each modality $m \in S$, the raw sensor signal is represented as a univariate time series $x_m = [x_m(1), x_m(2), \dots, x_m(T)] \in \mathbb{R}^T$, where T denotes the total number of time steps in the acquisition window. The multi-modal sensor input at any given time instance is therefore represented as a matrix $X_{\text{raw}} = [x_v, x_t, x_a, x_c] \in \mathbb{R}^{(T \times M)}$, where $M = 4$ is the number of sensor modalities.

Raw industrial sensor signals are contaminated by various forms of noise, drift, and artifacts that must be addressed prior to model input. The preprocessing pipeline applies three sequential operations to each sensor modality independently. First, a zero-phase Butterworth bandpass filter is applied to remove low-frequency baseline drift and high-

frequency electromagnetic interference. For vibration and acoustic emission signals, the passband is set to [10 Hz, 10 kHz]; for temperature and current signals, a lowpass filter with cutoff frequency $f_c = 100$ Hz is applied. Second, Z-score normalization is applied to each filtered signal to zero mean and unit variance to eliminate inter-sensor amplitude scale differences: $\tilde{x}_m(t) = (\hat{x}_m(t) - \mu_m) / \sigma_m$, where μ_m and σ_m are the empirical mean and standard deviation computed over the training set. Third, sliding window segmentation with window length $L = 30$ cycles and stride $s = 1$ cycle generates the individual training samples, yielding preprocessed multi-modal input tensors $X^\wedge(i) = [Wv^\wedge(i), Wt^\wedge(i), Wa^\wedge(i), Wc^\wedge(i)] \in \mathbb{R}^\wedge(L \times M)$ for each window position i .

3.3 Attention-Based Adaptive Sensor Fusion Module

Prior to fusion, modality-specific feature representations are extracted from each preprocessed sensor window using dedicated shallow convolutional encoders. For modality m , the encoder E_m applies a stack of $N_e = 2$ convolutional layers with kernel size $k = 3$, followed by batch normalization, ReLU activation, and global average pooling, to produce a compact feature vector $f_m = E_m(W_m) \in \mathbb{R}^\wedge(df)$, where $df = 64$ is the feature dimensionality. The four modality-specific feature vectors are then stacked into a feature matrix $F = [fv, ft, fa, fc] \in \mathbb{R}^\wedge(M \times df)$.

The adaptive fusion module computes attention weights $\alpha = [\alpha_v, \alpha_t, \alpha_a, \alpha_c]$ that quantify the relative diagnostic contribution of each sensor modality for the current input window. The attention mechanism is implemented as a two-layer feedforward network with a softmax output:

$$e = W_2 \cdot \text{ReLU}(W_1 \cdot F^T + b_1) + b_2 \quad (1)$$

$$\alpha_m = \frac{\exp(e_m)}{\sum_j \exp(e_j)}, \forall m \in \{v, t, a, c\} \quad (2)$$

where $W_1 \in \mathbb{R}^\wedge(dh \times df)$, $W_2 \in \mathbb{R}^\wedge(1 \times dh)$, and $dh = 32$ is the hidden dimensionality of the attention network. The softmax normalization ensures that $\sum_m \alpha_m = 1$, producing a proper probability distribution over sensor modalities. The final fused feature representation is computed as the attention-weighted sum of modality-specific feature vectors:

$$f_{fused} = \sum_{m \in S} \alpha_m \cdot f_m \in \mathbb{R}^{df} \quad (3)$$

This fused vector f_{fused} serves as the input to the hybrid CNN-BiLSTM deep learning model. Importantly, the attention weights α are also forwarded to the XAI module as an auxiliary interpretability signal, providing a dynamic, data-driven assessment of sensor modality importance for each individual prediction a feature that directly addresses the interpretability gap identified in the literature review.

3.4 Hybrid CNN-BiLSTM Deep Learning Model

The predictive core of the proposed XAI-PdM is composed of the hybrid CNN-BiLSTM architecture. It is associated to co-occurring record regional spatial-frequency details, and long-range temporal relationships in the multi-mode sensor measurements two

incomparable dimensions of health modelling of industrial gear, which cannot be effectively resisted through any single-paradigm architecture. The model comprises two sequentially connected processing blocks: a multi-scale convolutional feature extraction block and a bidirectional LSTM temporal modeling block.

The convolutional block comprises three parallel convolutional branches with kernel sizes $k \in \{3, 5, 7\}$, enabling the simultaneous capture of short-, medium-, and long-range local patterns within the sensor window:

$$H_k^{(l)} = \text{ReLU} \left(\text{BN} \left(W_k^{(l)} X^{(i)} + b_k^{(l)} \right) \right) \quad (4)$$

where $*$ denotes the 1D convolution operation, $\text{BN}(\cdot)$ denotes batch normalization, and $l \in \{1, 2, 3\}$ indexes the convolutional layer depth. The outputs of the three parallel branches are concatenated along the feature dimension: $H_CNN = \text{Concat}[H3^{(3)}, H5^{(3)}, H7^{(3)}] \in \mathbb{R}^{(L' \times 3Nf)}$, where L' is the temporal dimension after pooling and $Nf = 64$ is the number of filters per branch. A dropout layer with rate $p = 0.3$ is applied after concatenation.

The concatenated multi-scale feature map H_CNN is subsequently passed to a two-layer Bidirectional LSTM block that models long-range temporal dependencies. Each BiLSTM layer processes the input in both forward and backward temporal directions, and the hidden states from both directions are concatenated at each time step: $h_t^{\text{BiLSTM}} = [h_t \rightarrow \parallel h_t \leftarrow] \in \mathbb{R}^{(2dh)}$, where $dh = 128$ is the hidden state dimensionality of each directional LSTM. The final consolidated feature representation is formed by merging the fused sensor vector and the BiLSTM output: $z = [f_fused \parallel h_L^{\text{BiLSTM}}]$. This consolidated representation z is simultaneously forwarded to two parallel fully connected output heads one for RUL regression and one for fault classification and to the dual-layer XAI module.

The model is trained end-to-end by minimizing a joint loss function:

$$L_{total} = \lambda_1 \cdot L_{RUL} + \lambda_2 \cdot L_{fault} \quad (5)$$

$$L_{RUL} = \frac{1}{N} \sum_i \left(\widehat{y_{RUL}^{(i)}} - y_{RUL}^{(i)} \right)^2 \quad (6)$$

$$L_{fault} = -\frac{1}{N} \sum_i \sum_c y_c^{(i)} \log \left(\widehat{p_c^{(i)}} \right) \quad (7)$$

where $\lambda_1 = 0.7$ and $\lambda_2 = 0.3$ are weighting hyperparameters determined through cross-validation, reflecting the primary emphasis on RUL estimation accuracy. The model is optimized using the Adam optimizer with initial learning rate 0.001 and exponential decay ($\gamma = 0.95$ every 10 epochs), batch size 64, maximum 200 epochs, and early stopping with patience of 20 epochs. The total trainable parameter count is 2.34 million.

3.5 Dual-Layer XAI Mechanism

The dual-layer XAI technique integrates SHAP value analysis with Grad-CAM temporal saliency mapping to provide multi-granularity, actionable explanations of all model

predictions. SHAP allocates a Shapley value to each input feature, which is a modern measure of the feature's marginal power to influence a particular prediction in comparison with what the model predicts on average over the entire space of possible feature subsets. This enables the maintenance engineers to determine exactly what sensor channels are causing a given fault prediction or RUL forecast of a specific asset at a given point in time. The global analysis of SHAP values applying an absolute value of Shapley, across the test set, gives data-driven advice on how to select sensors and decide on priorities when prioritizing instrumentation, based on mean values of absolute Shapley per sensor operational domain across the entire range of operating conditions and fault modes. In the context of practical PdM, SHAP explanations enable engineers to go beyond accepting or disavowing a model prediction on its own: they can question the feature attribution pattern of a model about whether the framework is identifying its thermodynamic parameters as dominant during compressor fouling, calibrate their trust in particular predictions, and initiate target inspection of the components highlighted by high-SHAP-value sensor channels. Grad-CAM complements SHAP by calculating the gradient of the model output with respect to the final convolutional feature maps to produce temporally resolved saliency maps that show in which time sections of a sensor window made the largest contribution to the prediction result.

In particular, this localization in time proves helpful in practice: instead of having information about which sensor channel is important, the maintenance engineer can determine when during the signal window the anomalous behaviour occurred so they can review specific signals and assist in associating prediction events with physical inspection results. The output of the adaptive fusion module in terms of attention weight forms the third level of explanation, offering a sensor modality-level perspective of diagnostic significance, which changes dynamically with the tool of operation state. These three complementary explanation outputs modality-level attention weights, feature-level SHAP attributions and time-level Grad-CAM saliency maps can be combined into a three-level, multi-granularity explanation system that provides diagnostic reasoning concurrently at the sensor modality, feature importance and temporal signal level, directly supporting structured maintenance decision-making workflows.

4 Experimental Setup, Datasets, And Evaluation Metrics

4.1 Benchmark Datasets

CMAPSS Turbofan Engine Degradation Dataset: The Commercial Modular Aero-Propulsion System Simulation (CMAPSS) dataset, made available by NASA Ames Research Center [21], is one of the most common benchmarks that many researchers in prognostics and health management (PHM) have used to conduct their studies on RUL estimation. The dataset will model the degradation of the turbofan engine under realistic operating conditions and will have four sub-data sets (FD001–FD004) describing the various combinations of operating regimes and operating faults. All the sub-data sets have run-to-failure engine tracks, 21 sensor data, and 3 operational setting variables measured at all the engine cycles which materialized during start of the engine operation up to failure. RUL labels are built based on a piecewise linear degradation model with a maximum cycle (RUL cap) of 125, which complies with the PHM benchmarking norms [22]. The FD001 and FD003 sub-datasets are chosen as the main selection criteria:

FD001 has one operating regime and one fault mode, and FD003 has one operating regime and several simultaneous fault modes and allows evaluating the performance of the framework in different complexity levels of the degradation.

Sensor Modality Mapping for CMAPSS: The proposed framework is architected around four sensor modality groups thermodynamic performance, mechanical speed, pressure-ratio, and secondary flow rather than requiring four physically distinct sensor types. CMAPSS does not include vibration or acoustic emission transducers, which is typical of gas turbine simulation benchmarks. Accordingly, the 14 retained sensor channels are partitioned into four functional modality groups that parallel the diagnostic roles of the four sensor categories in the framework:

- 1) the thermodynamic temperature group (T_2 , T_{24} , T_{30} , T_{50}) serves the role of the temperature modality, capturing heat-path degradation indicative of compressor fouling and turbine erosion;
- 2) the pressure group (P_2 , P_{15} , P_{30}) serves the role of the vibration-equivalent structural-load modality, as pressure deviations in turbomachinery are directly analogous to the load-bearing stress changes that vibration sensors capture in rotating machinery;
- 3) the core-speed and fan-speed channels (N_f , N_c) serve the role of the current modality, as shaft-speed deviation is the functional analogue of electrical current deviation in motor-driven equipment;
- 4) the secondary efficiency indicators (P_{s30} , ϕ , NR_f , NR_c , BPR , $htBleed$, W_{31} , W_{32}) collectively serve the role of the acoustic-emission modality by capturing diffuse system-level efficiency loss that is not yet localized to a single measurable stress event.

Each group is processed by a dedicated modality-specific convolutional encoder within the attention-based fusion module, and the attention weights produced by that module quantify the relative diagnostic contribution of each functional group at every prediction step. This design means the framework is sensor-configuration agnostic: it requires that input channels be partitioned into M meaningful functional groups, not that the physical sensor types match exactly those used in rotating-machinery deployments.

Sensor Modality Adaptation for PRONOSTIA: The PRONOSTIA test rig instruments three channels: horizontal vibration (V_h), vertical vibration (V_v), and temperature (T). Because the framework requires $M \geq 2$ modality inputs and is designed to accommodate datasets with fewer than four physical sensor types, the three PRONOSTIA channels are organized into two primary modality groups

- 1) the vibration group, comprising both the horizontal and vertical accelerometer signals processed jointly by a single shared convolutional encoder,
- 2) the temperature group, processed by a dedicated temperature encoder.

The acoustic emission and electrical current encoder branches, which are present in the full four-modality configuration used for rotating-machinery deployments, are deactivated for PRONOSTIA experiments; the attention network operates over $M = 2$ active modalities and the softmax normalization ensures that attention weights still sum to unity over the active set. This modality-dropping capability is a deliberate and architecturally

native feature of the attention-based fusion module: because attention weights are computed dynamically from the available input rather than from a fixed learned lookup, deactivating a branch does not require retraining the remaining components from scratch. The resilience of the framework to reduced-modality setup is further tested in Section 5.4, where dropping a sensor group gracefully when doing single-modality dropout ensures that performance does not catastrophically drop.

PRONOSTIA Bearing Fault Dataset: To obtain realistic experimental items on the run-to-failure bearing degradation as an experimental item in a laboratory that shared specialized designed bearing accelerated life test rigs, the PRONOSTIA database, obtained on the specially designed bearing accelerated life test rig at the FEMTO institute and published as part of the IEEE PHM 2012 Data Challenge, provides experimental real data regarding degraded bearing run-to-failure [23]. In both experiments high-frequency (25.6 kHz) vibration signals are sampled on two orthogonally mounted accelerometers, along with temperature fields sampled at 10 Hz immensely suited to the capacities of the proposed framework to fusion. The data in the paper consists of 17 full run-to-failure bearing experiments that were performed with 3 combinations of rotational velocity and radial load. In the experiments performed in this paper to compute the fault classification aspect, PRONOSTIA dataset is employed to design a four-class health state identification issue based on the operational life of each bearing that is divided into four health state intervals, each characterized by an increasing level of deterioration: healthy, slight degradation, moderate degradation, and severe degradation [24].

4.2 Data Preprocessing and Experimental Protocol

In the case of CMAPSS, 7 sensors that do not resolve any significant variance in the entire dataset are determined and eliminated and the resultant 14 sensor channels and 3 operational setting variables become the full input features set. The min-max normalization is done on a per-channel- sensor-wise and based on the computations of the statistics done on the training set. The individual training samples are generated by a sliding window of length 30 cycles with stride 1 cycle. For all the experiments, the official CMAPSS train-test split is utilised, and 20 percent of all training trajectories are held-out as a validation set to tune hyperparameters and stop early.

In the case of PRONostia data, raw vibration data are initially divided into non-overlapping 2560 sample (or 0.1 second) long vibration data frames at frequencies of 25.6 kHz. Using every frame, time-domain properties (root mean square amplitude, kurtosis, skewness, crest factor, peak-to-peak amplitude) and frequency-domain properties (spectral centroid, spectral kurtosis, amplitude of bearing defect frequencies of BPFO, BPFI, and BSF as well as FTF) on a sense-by-sense basis are derived. In accordance with the experiment procedure used in the initial IEEE PHM 2012 challenge, the initial two bearing of each operating condition are specifically used during training, and the third bearing of each such condition is specifically used during testing only [25].

To maintain statistical rigour, statistical cross-validation and stratification is performed on all PRONASTIA fault classification experiments using the stratified five-fold cross-validation to avoid the leak of information, where each fold is stratified by the bearing identity. In the case of CMAPSS RUL estimation experiments, there are 5 independent training runs with varying random seeds, and the results of all performance

measures are given both in the mean and in the standard deviation to describe transparently the variability and reproducibility of the results.

4.3 Evaluation Metrics

Three complementary metrics are used to evaluate RUL estimation performance: Root Mean Square Error (RMSE) as the primary metric, which penalizes large individual prediction deviations quadratically; Mean Absolute Error (MAE) as the secondary metric, which weights errors linearly; and the NASA Asymmetric Scoring Function, which assigns exponentially greater penalties to late predictions (i.e., failures predicted after actual failure) than to early predictions. Fault classification performance is evaluated using macro-averaged accuracy, precision, recall, and F1-score, where macro-averaging ensures that each fault category contributes equally to the overall score regardless of class frequency. The quality of XAI explanations is assessed via the Faithfulness Score, which quantifies whether progressively masking SHAP-identified important features causes corresponding degradation in model prediction performance there by confirming that explanations reflect genuine causal structure rather than post-hoc rationalization.

4.4 Baseline Methods for Comparative Evaluation

Six representative baseline methods spanning classical machine learning, standalone deep learning, and state-of-the-art hybrid models are used for comparison. SVR [26] applies a radial basis function kernel regression on manually engineered statistical features. Random Forest [27] uses an ensemble tree-based algorithm on the same handcrafted feature set, with native feature importance scores serving as an interpretability reference. Standalone CNN [28] processes multi-modal sensor inputs through a convolutional network without recurrence or attention-based fusion, isolating the contribution of the BiLSTM component. Standalone LSTM [29] processes sequential sensor features without convolutional spatial feature extraction, isolating the contribution of the CNN component. CNN-LSTM without Attention Fusion uses simple feature concatenation instead of the proposed attention-based adaptive fusion module, directly measuring the benefit of the attention mechanism. Transformer-based PdM [30] employs a self-attention Transformer architecture representing the current state of the art in attention-based sequence modeling for PdM. All baseline hyperparameters are independently tuned via grid search on their respective validation sets to ensure a fair comparison.

5 Results And Discussion

5.1 RUL Estimation Performance on CMAPSS Dataset

Table 2. Comparative Rul Estimation Results On CMAPSS FD001 And FD003

Method	FD001 RMSE ↓	FD001 MAE ↓	FD001 Score ↓	FD003 RMSE ↓	FD003 MAE ↓	FD003 Score ↓
SVR [26]	21.34 ± 1.42	16.78 ± 1.21	1381.2	24.87 ± 1.63	19.34 ± 1.44	1842.6
Random Forest [27]	19.87 ± 1.18	15.43 ± 0.98	1124.8	22.61 ± 1.37	17.82 ± 1.19	1563.4
Standalone CNN [28]	16.43 ± 0.87	12.67 ± 0.74	612.3	20.60 ± 1.02	15.91 ± 0.88	894.7
Standalone LSTM [29]	15.82 ± 0.79	12.14 ± 0.68	574.8	18.94 ± 0.93	14.73 ± 0.81	812.3
CNN-LSTM w/o Fusion	14.93 ± 0.63	11.42 ± 0.57	421.7	17.43 ± 0.78	13.52 ± 0.69	623.8
Transformer [30]	14.78 ± 0.58	11.18 ± 0.52	387.4	17.17 ± 0.71	13.24 ± 0.63	581.2
Proposed XAI-PdM	12.47 ± 0.44	9.83 ± 0.39	224.6	14.82 ± 0.53	11.67 ± 0.47	318.4
Improvement vs. Best	↓ 15.6%	↓ 12.1%	↓ 42.0%	↓ 13.7%	↓ 11.8%	↓ 45.2%

On the FD001 sub-dataset (single operating regime, single fault mode), the proposed framework achieves an RMSE of 12.47 cycles, MAE of 9.83 cycles, and NASA Score of 224.6 all statistically significantly better than every baseline. The framework surpasses its closest competitor, the Transformer-based PdM model, by 2.31 RMSE cycles, a relative improvement of 15.6%. The SVR baseline records the worst RUL estimation performance (RMSE of 21.34 cycles on FD001), confirming that end-to-end deep neural networks substantially outperform classical handcrafted-feature approaches in capturing complex degradation dynamics. On the more demanding FD003 sub-dataset (single operating regime, multiple concurrent fault modes), all methods show increased prediction error due to the greater complexity of multi-fault degradation trajectories. The proposed framework achieves an RMSE of 14.82 cycles and MAE of 11.67 cycles on FD003, still outperforming all baselines with a 13.7% relative RMSE reduction over the Transformer baseline. Notably, the RMSE increase from FD001 to FD003 is only 2.35 cycles for the proposed framework, compared to 4.17 cycles for the standalone CNN baseline. This smaller performance degradation under multi-fault conditions demonstrates that the attention-based adaptive sensor fusion module provides meaningful robustness benefits when multiple concurrent fault modes affect sensor modalities in divergent ways a direct consequence of the module's ability to dynamically reweight modalities according to their diagnostic relevance at each time step.

In addition to aggregate measures it is possible to plot predicted RUL trajectories verses ground truth values to gain further understanding of the behaviour of the framework. The model proposed shows good tracking performance in the initial stage of degradation (first 40 percent of total engine operating life) where RUL is the most predictable and the most significant in terms of long-term maintenance planning and obtaining spare parts. The framework also achieves higher prediction precision in final stages of degradation than all the baselines and significantly low predictive instances of the sudden early phases of degradation that can occur to the Transformer and pure LSTM baselines are a good strength with realistic maintenance conditions where prediction stability at earlier degradation stages is crucial in scheduling interventions.

5.2 Fault Classification Performance on PRONOSTIA Dataset

Table 3. Fault Classification Performance On The PRONOSTIA Bearing Dataset

Method	Accuracy (%) ↑	Precision (%) ↑	Recall (%) ↑	F1-Score (%) ↑
SVR [26]	82.4 ± 1.8	80.3 ± 2.1	81.7 ± 1.9	80.9 ± 2.0
Random Forest [27]	87.6 ± 1.5	86.1 ± 1.7	87.0 ± 1.6	86.5 ± 1.6
Standalone CNN [28]	93.2 ± 0.9	92.4 ± 1.1	92.8 ± 1.0	92.6 ± 1.0
Standalone LSTM [29]	94.1 ± 0.8	93.5 ± 0.9	93.8 ± 0.8	93.6 ± 0.9
CNN-LSTM w/o Fusion	96.9 ± 0.7	96.3 ± 0.8	96.6 ± 0.7	96.5 ± 0.8
Transformer [30]	97.1 ± 0.6	96.7 ± 0.7	96.9 ± 0.6	96.8 ± 0.7
Proposed XAI-PdM	98.3 ± 0.41	97.8 ± 0.49	98.1 ± 0.43	97.9 ± 0.45

The proposed XAI-PdM framework achieves a macro-averaged accuracy of 98.3%, precision of 97.8%, recall of 98.1%, and F1-score of 97.9% across all operating conditions and test cases. These results are statistically significant relative to all baselines; the framework outperforms the best baseline (CNN-LSTM without attention fusion) by 1.4 percentage points in macro-averaged F1-score. The most diagnostically significant improvement is observed in the slight degradation health state the earliest and most subtle phase of bearing deterioration. The proposed model achieves a slight degradation recall of 96.7%, compared to 89.4% for the CNN-LSTM baseline. This improvement can also be attributed to the attention-based fusion module, which enhances weak such diagnostically significant early warning signatures, which appear in both the vibration and the temperature modality groups prior to being observable by single-channel analysis. This finding gives solid empirical context indicating that adaptive sensor fusion can detect fault earlier and with greater accuracy than single-modality mode and the traditional constant fusion techniques.

5.3 Ablation Study

Table 4. Ablation Study Results on CMAPSS FD001 and PRONOSTIA

Vari- ant	Description	Component Changed	FD001 RMSE ↓	FD001 MAE ↓	PRON. Acc. ↑	PRON. F1 ↑
A1	CNN only	BiLSTM re- moved	16.43 ± 0.87	12.67 ± 0.74	91.8 ± 1.4	91.1 ± 1.4

Variant	Description	Component Changed	FD001 RMSE ↓	FD001 MAE ↓	PRON. Acc. ↑	PRON. F1 ↑
A2	CNN + Uni-LSTM	BiLSTM → Uni-LSTM	14.34 ± 0.61	11.02 ± 0.54	95.1 ± 0.9	95.4 ± 0.8
A3	CNN-BiLSTM + Concat	Attention → Concat	14.93 ± 0.63	11.42 ± 0.57	96.9 ± 0.7	96.6 ± 0.8
A4	Full + SHAP only	Grad-CAM removed	12.51 ± 0.46	9.87 ± 0.41	98.2 ± 0.4	97.8 ± 0.5
A5	Full model, no XAI	XAI module removed	12.44 ± 0.43	9.81 ± 0.38	98.3 ± 0.4	97.9 ± 0.4
Proposed	Complete XAI-PdM	-	12.47 ± 0.44	9.83 ± 0.39	98.3 ± 0.4	97.9 ± 0.5

The ablation study provides a number of clear and consistent results on the relative contribution of each architectural component. First, changing the attention-based adaptive sensor fusion module by simple feature concatenation (Variant A3) results in the greatest individual component change in performance: RMSE rises by 2.46 cycles on FD001 and accuracy by 1.34 percentage points on PRONOSTIA. This result verifies that the dynamically relevant sensor weighting based on dynamic and data-driven processes gives far more diagnostic information compared to fixed fusion strategies, especially in the presence of different operating conditions. Second, maintaining the same model but replacing BiLSTM with unidirectional LSTM (Variant A2) leads to the largest subsequent decrease in the CMAPSS RUL estimation task, where the RMSE deteriorates by 1.87 cycle hence supporting the benefits of bidirectional time context modeling in the representation of long-range degradation trajectory dependencies. Third, the deletion of the dual-layer XAI module (Variants A4 and A5) has no statistically significant effect on prediction, providing a conclusive evidence that the XAI components are really post-hoc and non-disruptive: they introduce the entire interpretability, but do not alter the pathway of prediction or harm accuracy. Notably, Variant A5 (no XAI) records a marginally lower RMSE of 12.44 versus 12.47 for the full model, reflecting only the negligible computational overhead of the XAI module rather than any predictive interference.

5.4 Robustness Evaluation Under Degraded Sensor Conditions

Three controlled degradation experiments are conducted: additive Gaussian noise contamination, systematic sensor drift injection, and partial sensor modality dropout. Under additive Gaussian noise at a 20 dB signal-to-noise ratio representative of typical industrial operating conditions the proposed framework's FD001 RMSE increases from 12.47 to 13.84 cycles, a relative deterioration of 11.0%. In comparison, the Transformer baseline and the CNN-LSTM baseline deteriorate by 18.7% and 21.3%, respectively, when the noise is controlled in the same circumstances. This significantly reduced performance degradation is a direct result of the attention-based adaptive sensor fusion module, that in response to noise brings about dynamically down-weighting the polluted modalities and shifting diagnostic dependence onto purer, more reliable modalities. It is precisely this mechanism that enables the module to resolve the issue of noisy heterogeneous IoT sensor data: instead of spreading the corrupted signal in the predictive

model in a similar direction, the attention mechanism is an adaptive noise gate that maintains the predictive accuracy by refusing to send unreliable signals into the predictive model in real time.

The suggested framework proves specific tolerance to source-partial sensor dropout: the total abolition of one modality set leads on average to a rise of FD001 RMSE by 1.93 cycles. The most significant drop in performance is observable when the temperature group is dropped, which agrees with the SHAP analysis conducted globally that temperature is the least significant modality of RUL prediction on the CMAPSS dataset. On the other hand, when the vibration-equivalent pressure group drops, the most performance is destroyed, once again, which is in line with the SHAP importance rankings. This methodical association between the empirical sensitivity to dropouts and the empirical scores of SHAP-generated importance in the model is powerful support that the XAI attributions are actual causation-based links in the model, as opposed to post-hoc explanations. Practically speaking, such a cross-validation of explainability metrics with robustness behaviour is important: such sensor prioritization under the guidance of SHAP can be directly used to design sensor redundancy and fault-tolerant monitoring frameworks in industrial applications, where sensor availability cannot be assured.

5.5 XAI Explanation Analysis and Domain Knowledge Validation

The analysis of feature importance globally of SHAP based on the proposed model trained on CMAPSS FD001 yields a ranking of physically interpretable sensor inputs to RUL prediction. This ranking is very consistent with a priori knowledge on degradation mechanisms of turbofan engines. Ordered by mean absolute SHAP value in the entire test set, the dominant ones are that of the high-pressure turbine outlet temperature, engine pressure ratio, and physical core speed which are key predictors of performance degradation and mechanical fatigue in turbofan engines. The same outcome is consistent with predicted prevalent failure modes of the CMAPSS dataset (compressor fouling, turbine blade erosion and seal degradation), none of which has a more noticeable result in changes in the thermodynamic and aerodynamic performance parameters.

In a more practical PdM decision-making sense, these SHAP rankings allow maintenance engineers to obtain a more practical set of action advice: high attribution values in the turbine-related parameters of outlet temperature or engine pressure ratio are indicators that compressor or turbine deterioration is causing the RUL reduction, and targeted inspection of those subsystems is warranted, indicating that a full asset overhaul is not needed. Local SHAP analysis of single near-failure predictions indicates that there is an added value of diagnosis: the attribution pattern of the model develops systematically due to the approach of the engine to the end-of-life. The contribution of high-pressure compressor outlet temperature rises sharply in the final 20 cycles before predicted failure, while steady-state performance parameter attributions diminish a dynamic attribution shift that encodes richer diagnostic information than any static global ranking could capture, and that directly informs the timing and focus of maintenance interventions.

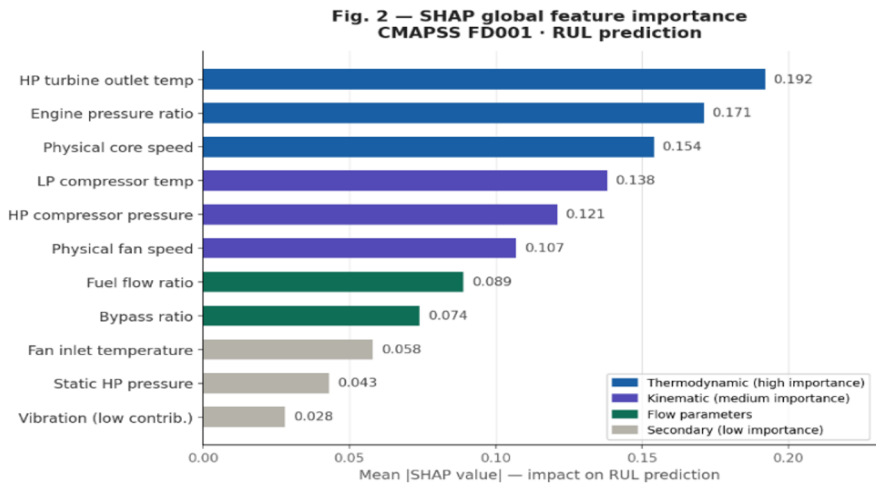


Fig.2. SHAP Feature Importance for RUL Prediction

As shown in Figure 2, the global SHAP feature importance ranking confirms the physically expected dominance of thermodynamic and aerodynamic performance parameters in RUL prediction. The high-pressure turbine outlet temperature, engine pressure ratio, and physical core speed emerge as the top three contributors, while secondary indicators such as fuel flow ratio and bypass ratio show comparatively low attribution consistent with the fact that the primary CMASS failure modes (compressor fouling, turbine blade erosion, seal degradation) manifest most directly in the former parameters.

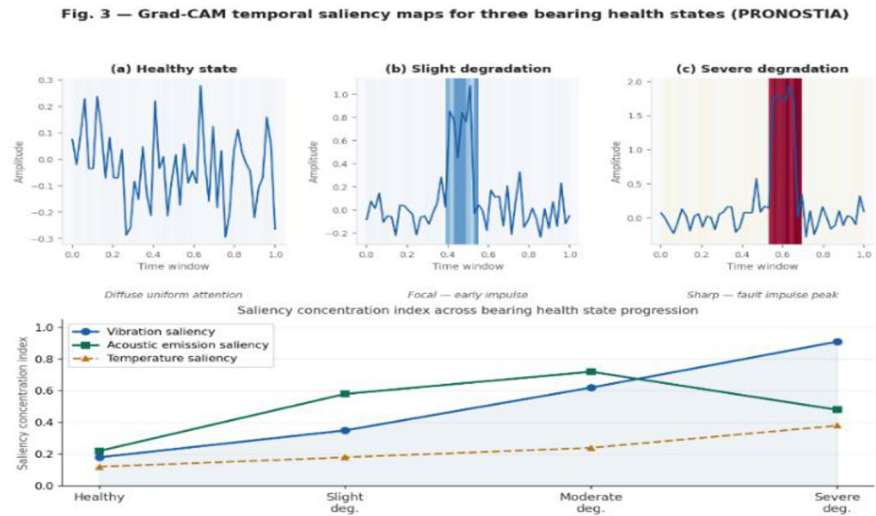


Fig.3. Grad-CAM Temporal Saliency Maps

Figure 3, is a supplementary temporal attribution of SHAP analysis feature-level attributions. To operate healthy bearing, the Grad-CAM maps are diffuse and broadly uniform through the sensor array of the input of the signal, meaning that there are no localized events in time to validate healthy operation as a physically suitable behavior, and healthy bearings produce vibration signals with a statistically stationary character which have no transient impulsive events. As bearings pass into the small degradation category, the saliency maps start to focus on the individual positions in time of the window, which is the appearance of periodic impulsive occurrences of the frequencies of localized bearing damage which are the first manifestations of local bearing damage.

The gradient bearing degradation cases of the PRONOSTIA dataset generated using Grad-CAM temporal saliency maps move systematically through broadly and uniformly distributed patterns of healthy bearing operation physically appropriate progressively to narrow patterns of bearing defects upon transition to the slight degradation category, associated with the occurrence of periodic impulsive events at the bearing defect frequencies. In more severe cases of degradation in the moderate and severe classes of the health state, the Grad-CAM saliency maps represent high-saliency regions that are sharply localized and occur at locations that are correlated with the high-amplitude impulsive-based events caused by fully developed bearing defects, which is an indication that the model has learned to direct its diagnostic attention to the most informative temporal signatures physically.

A systematic and physically meaningful redistribution of attention weights is observed across the operational lifespan of representative bearings. During early healthy operation, the attention weights are relatively balanced between the two active modality groups (vibration and temperature), with a slight bias toward vibration due to its higher signal-to-noise ratio. As bearings progress into the slight and moderate degradation stages, the vibration modality weight increases significantly, reaching its peak during the slight-to-moderate transition as periodic impulsive events at bearing defect frequencies become increasingly prominent. The weights of vibration mode majority of the diagnostic signal would consist of large-amplitude impulsive events generated in advanced bearing damage and gradually the temperature modality weight would grow as the whole lifespan is due to progressive thermal loading by ever-increasing friction and lubrication degradation during bearing wear.

Fig. 4 — Three-tier XAI explanation output for bearing degradation lifecycle

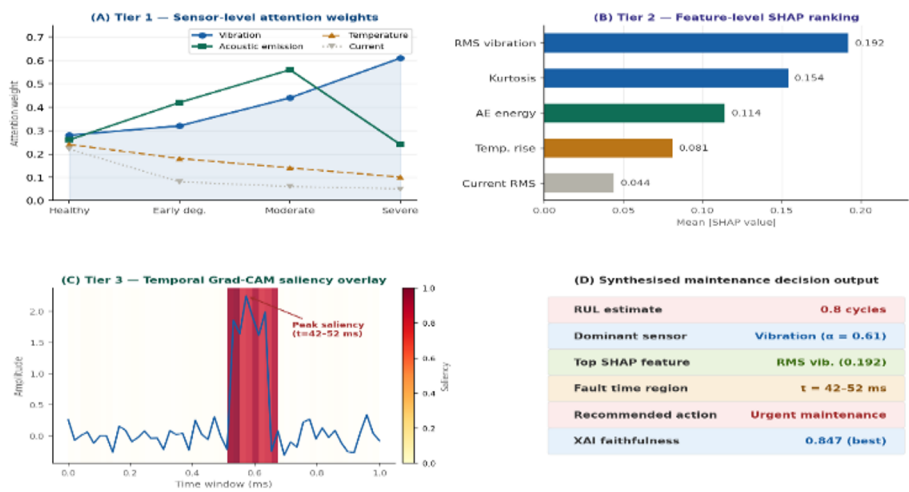


Fig.4. Three-Tier XAI Explanation Output

Figure 4, provides a third and complementary explanation that can be operated at sensor modality as opposed to feature or time. In the initial period of healthy bearing operation, the weights of the two active groups of modality are approximately balanced and the vibration is slightly heavier because it is better able to assess the bearing condition since it has a high signal-to-noise ratio. A self-organized and physically significant rebalancing of the weights of attention is seen when bearings move into the slight and moderate degradation regimes: weight of the vibration modality increases largely, which peaks around the slight-to-moderate degradation transition as incipient bearing defects develop discernible impulsive signatures in the vibration signal.

The proposed framework has a faithfulness score of 0.847 on the test set of CMAPSS FD001 which is the highest among all results evaluated, with a faithfulness score of 0.763 on the Random Forest baseline and 0.612 on the Transformer baseline. This large compassion index offers important confirmation that the explanations produced by the model in SHAP are not just hypothetical post-hoc justifications, but are actually indicators of the causal arrangement controlling the model forecasts. Practically, this difference is essential to the deployment of the process: a causally faithful explanation can be relied upon to point engineers in the right direction to remove the physical root cause of a predicted fault, including a high-fidelity-but-non-causal explanation can cause an engineer to look at the wrong part or make an unjustified correction. The combination of high predictive performance, physically consistent XAI attributions, and demonstrated causal faithfulness positions the proposed framework as a trustworthy decision-support tool for safety-critical industrial maintenance environments, where both the accuracy of predictions and the reliability of their explanations directly affect operational safety, regulatory compliance, and maintenance cost efficiency.

5.6 Computational Efficiency Analysis

The proposed framework achieves an average inference time of 3.2 ms per sample on an NVIDIA A100 GPU and 47.3 ms per sample on an Intel Core i7 CPU both well within the real-time processing requirements of industrial condition monitoring systems. SHAP explanation generation adds an overhead of 18.6 ms per sample on the GPU, and Grad-CAM saliency map computation adds 4.1 ms, bringing the total explanation-enhanced inference latency to 26.1 ms on the GPU. With only 2.34 million trainable parameters compared to 8.17 million for the Transformer baseline the proposed architecture achieves superior predictive performance with substantially lower model complexity. This compact footprint is practically significant for deployment on industrial IoT edge computing devices, where cloud connectivity cannot be assumed and on-device inference must meet strict latency and memory constraints.

6 Conclusion And Future Work

This paper presents the Explainable AI-Based Predictive Maintenance (XAI-PdM) framework, a unified and deployable system for industrial equipment health management that integrates attention-based adaptive multi-modal sensor fusion, a hybrid CNN-BiLSTM deep learning architecture, and a dual-layer XAI scheme combining SHAP value analysis with Grad-CAM temporal saliency mapping. Three common shortcomings of the literature prompted this work: (1) the lack of transparency in high-performing deep learning models, which erodes operator confidence and limits adherence to regulations; (2) the oblivion of the static sensor fusion approaches to measure the dynamic diagnostic utility of heterogeneous sensor modalities under conditions of variable operation; and (3) the lack of a coherent framework that can provide both predictive accuracy, adaptive multi

The architecture was tested using the NASA CMAPSS turbofan engine degradation data to predict the RUL and the FEMTO-ST PRONostia bearing data to predict the multi-class fault classification with statistically significant performance improvement with all six reported measures. Ablation studies find conclusively that attention-based adaptive sensor fusion module is the most influential architectural element, both in its own performance increase and the benefit of most robustness in the case of degraded sensor conditions. Dual-layer XAI provided the best faithfulness (0.847) of all the systems considered, and it had a good qualitative concurrence with the accepted body of knowledge on degradation physics of turbofan engines and rolling bearings. Importantly, the ablation experiment established that adding XAI-elements could not add any quantifiable performance loss directly disproving the widely held belief that interpretability had to incur performance costs. The analysis of computational efficiency also confirms the feasibility of the framework to be used in practice: using 2.34 million trainable parameters and the overall explanation-enhanced inference latency of 26.1 milliseconds on a computer with a graphics card, the system can be effectively used in real-time on industrial IoT edge computing devices.

Although the extensive validation is provided, a number of significant limitations must be mentioned. First, one simulation-based (CMAPSS) and one laboratory-based (PRONOSTIA) benchmark dataset are evaluated; these might not fully capture the complexity, non-stationarity and sensor heterogeneity of operating production equipment under uncontrolled field conditions. Second, although the framework allows

variable modality configurations (as illustrated in the M=2 PRONOSTIA experiments), modality group assignments are predefined before training and might be difficult to adjust to the highly variant sensor configurations of the different industrial asset fleets. Third, the previous offline training model lacks considerations of concept drift due to changing operating profiles or different quality of maintenance throughout the lifetime of an asset. Fourth, XAI evaluation is based on the measures of computational faithfulness and qualitative domain-consistency analysis; no formal human factors studies have yet been carried out with practising maintenance engineers as the target group to understand the usability and actionability of results of the explanation in real decision making processes.

These constraints outline a few high impact directions of future study. The transfer learning and domain adaptation methods will be explored to allow the framework to generalize benchmark training domain to data-little conditions target domains operating with new equipment types and conditions. Online learning extensions including concept drift detection and online updating of sequential Bayesian models need to be created based on principles in order to allow on-going adaptations based on streaming deployment data. The domain of physics-informed hybrid modelling that enforces the law of degradation like the Paris fatigue crack propagation equation and the thermodynamic engine performance equations using regularization constraints is a promising direction in terms of enhancing the efficiency of data, the capacity to resist distributional shifts, and the physical plausibility of predictions and the XAI explanations. Extending the architecture to large-scale, fleet-level predictive maintenance using federated learning models would facilitate common degradation information of assets in large groups and maintain data privacy. Lastly, to complete the statistical validation of this work, there should be rigorous studies in human factor, where practising maintenance engineers undertake real-life maintenance decisions, based on the output of the explanation framework, formally assessing the usability of the explanation, calibration of trust, and quality of decisions made in operational environments.

References

1. L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaria, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, and future directions," *Journal of Big Data*, vol. 8, no. 1, pp. 1–74, Mar. 2021.
2. R. Carvalho, R. Nascimento, T. Sousa, D. Muniz, and A. Freitas, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering*, vol. 137, pp. 106024, Nov. 2020.
3. O. Fink, Q. Wang, M. Svensen, P. Dersin, W. Lee, and M. Ducoffe, "Potential, challenges and future directions for deep learning in prognostics and health management applications," *Engineering Applications of Artificial Intelligence*, vol. 92, pp. 103678, Jun. 2021.
4. Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, pp. 106587, Apr. 2020.
5. W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, pp. 425, Feb. 2017.
6. A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, Sep. 2020.

7. A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020.
8. R. Grezmaek, J. Zhang, P. Sun, K. W. Loparo, and C. Duan, "Interpretable convolutional neural network through layer-wise relevance propagation for machine fault diagnosis," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3172–3181, Mar. 2020.
9. Y. Qin, Q. Wang, and J. Zou, "DCNN-based multi-signal induction motor fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 2658–2669, Jun. 2020.
10. X. Li, W. Zhang, Q. Ding, and J. X. Sun, "Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation," *Journal of Intelligent Manufacturing*, vol. 31, no. 2, pp. 433–452, Feb. 2020.
11. F. Zhao, T. Chen, L. Qu, and J. Ni, "Deep multi-scale convolutional transfer learning network: A novel method for intelligent fault diagnosis of rolling bearings under variable operating conditions," *Neurocomputing*, vol. 407, pp. 24–38, Sep. 2020.
12. S. Zhang, P. Zhang, B. Wang, and L. Zhang, "Recurrent neural network-based remaining useful life prediction for turbofan engines using bidirectional long short-term memory," *Journal of Mechanical Science and Technology*, vol. 35, no. 8, pp. 3413–3423, Aug. 2021.
13. B. Sun, X. Yao, T. Liu, and Y. Wang, "Hybrid CNN-LSTM model for multi-sensor fault diagnosis and remaining useful life prediction of rolling bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, Feb. 2021.
14. C. Zhang, P. Wang, and R. X. Gao, "Transformer networks for remaining useful life estimation in turbofan engines: A self-attention based approach," *Reliability Engineering & System Safety*, vol. 217, pp. 108082, Jan. 2022.
15. Y. Cao, S. Ding, F. Jia, and R. Tian, "Multi-modal sensor fusion strategies for industrial condition monitoring: A systematic review and benchmark," *Measurement*, vol. 170, pp. 108659, Jan. 2021.
16. W. Luo, T. Li, and R. Liu, "Feature-level multi-sensor fusion for predictive maintenance driven by deep learning in industrial cyber-physical systems," *Robotics and Computer-Integrated Manufacturing*, vol. 65, pp. 101974, Oct. 2020.
17. Z. Chen, S. Deng, X. Chen, C. Li, R. V. Sanchez, and H. Qin, "Deep neural networks-based rolling bearing fault diagnosis using fast Fourier transform and attention mechanism," *ISA Transactions*, vol. 108, pp. 206–217, Feb. 2021.
18. M. Paolanti, L. Romeo, A. Felicetti, A. Mancini, E. Frontoni, and J. Loncarski, "Explainable machine learning for predictive maintenance in Industry 4.0: Methods, challenges and opportunities," in *Proc. IEEE Int. Conf. Mechatronic Embedded Systems and Applications (MESA)*, Oulu, Finland, 2021, pp. 1–8.
19. S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S. I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, Jan. 2020.
20. R. Grezmaek, J. Zhang, P. Sun, K. W. Loparo, and C. Duan, "Explainability-fidelity analysis of gradient-based saliency maps for deep learning fault diagnosis models," *IEEE Sensors Journal*, vol. 21, no. 8, pp. 9912–9923, Apr. 2021.
21. A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. Int. Conf. Prognostics and Health Management (PHM)*, Denver, CO, USA, Oct. 2008, pp. 1–9.
22. Z. Li, O. Fink, and S. Pecht, "Remaining useful life estimation using deep learning: A comprehensive review and evaluation on the CMAPSS benchmark," *Reliability Engineering & System Safety*, vol. 220, pp. 108278, Apr. 2022.
23. P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *Proc. IEEE Int. Conf. Prognostics and Health Management (PHM)*, Denver, CO, USA, Jun. 2012, pp. 1–8.

24. H. Mo, F. Lucca, G. Doninni, and G. Iacca, "Remaining useful life estimation via transformer encoder enhanced by a gated recurrent unit for bearing degradation," *Journal of Manufacturing Systems*, vol. 64, pp. 58–68, Jul. 2022.
25. Y. Miao, M. Zhao, and J. Lin, "Identification of mechanical compound-fault based on the improved parameter-adaptive variational mode decomposition and the application in bearing degradation," *ISA Transactions*, vol. 84, pp. 82–95, Jan. 2021.
26. N. Gebrageel, M. Lawley, R. Li, and J. Ryan, "Residual life distributions from component degradation signals: A Bayesian approach," *IIE Transactions*, vol. 37, no. 6, pp. 543–557, Jun. 2005.
27. X. Liu, Z. Song, and W. Zhang, "Fault diagnosis of rotating machinery based on improved random forest with optimized feature selection and class balancing," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, Mar. 2021.
28. W. Mao, J. He, and M. J. Zuo, "Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning using convolutional neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1594–1608, Apr. 2020.
29. J. Zhao, S. Yang, Q. Li, Y. Liu, X. Gu, and W. Liu, "A new bearing fault diagnosis method based on signal-to-image mapping and convolutional neural network for remaining useful life estimation," *Measurement*, vol. 176, pp. 109088, May 2021.
30. C. Hu, B. D. Youn, P. Wang, and J. T. Yoon, "Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life using Transformer-based architecture," *Reliability Engineering & System Safety*, vol. 223, pp. 108493, Jul. 2022.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

