Proceedings of the 8th European Conference of the Prognostics and Health Management Society 2024 - ISBN – 978-1-936263-40-0

Integrating Network Theory and SHAP Analysis for Enhanced RUL Prediction in Aeronautics

Yazan Alomari[1], Marcia L. Baptista [2], and Mátyás Andó[3]

[1,3]*Eötvös Loránd University, Faculty of Informatics, Institute of Computer Science, Budapest, 1053, Hungary*
yazan@inf.elte.hu
am@inf.elte.hu

[2]*Delft University of Technology, Faculty of Aerospace Engineering, GB Delft, 2600, The Netherlands*
m.lbaptista@tudelft.nl

ABSTRACT

The prediction of Remaining Useful Life (RUL) in aerospace engines is a challenge due to the complexity of these systems and the often-opaque nature of machine learning models. This opaqueness complicates the usability of predictions in scenarios where transparency is crucial for safety and operational decision-making. Our research introduces the machine learning framework that significantly improves both the **interpretability** and **accuracy** of RUL predictions. This framework incorporates SHapley Additive exPlanations (SHAP) with a surrogate model and Network Theory to clarify the decision-making processes in complex predictive models and enhance the understanding of the hidden pattern of features interaction. We developed a Feature Interaction Network (FIN) that uses SHAP values for node sizing and SHAP interaction values for edge weighting, offering detailed insights into the interdependencies among features that affect RUL predictions. Our approach was tested across 44 engines, showing RMSE values between 2 and 17 and NASA Scores from 0.2 to 1.5, indicating an increase in prediction accuracy. Furthermore, regarding interpretability the application of our FIN, revealed significant interactions among corrective speed and critical temperature points key factors in engine efficiency and performance.

1. INTRODUCTION

In the interdisciplinary domain of Prognostics and Health Management (PHM), the accurate prediction of Remaining Useful Life (RUL) for industrial assets has become paramount (Ren et al., 2023). As aerospace, automotive, and manufacturing sectors increasingly depend on the reliability of their machinery, accurately predicting maintenance needs has become essential for ensuring safety, maximizing efficiency, and reducing costs. This necessity has driven the shift from traditional prognostic methods to advanced machine learning techniques (Calabrese et al., 2020; Deutsch & He, 2018). These modern methods utilize large datasets to effectively identify complex patterns and trends in machinery wear and tear, significantly enhancing our ability to predict equipment failures (Duc Nguyen et al., 2019).

However, the application of ML in PHM is limited by significant challenges, such as models interpretability. The "black box" nature of many ML algorithms, particularly those based on deep learning, obscures the decision-making processes underlying their predictions. This opacity is a considerable concern in fields when understanding the 'why' behind a prediction is as critical as the prediction itself, necessitating models that stakeholders can trust and interpret (Baptista et al., 2022; Kononov et al., 2023; Vollert et al., 2021).

Historical reliance on reliability and physics based models for RUL estimation, though effective, often staggers upon the complexities inherent in real-world operational scenarios. These traditional methods necessitate detailed domain knowledge and often lack the flexibility to adapt to different types of machinery (X. Li et al., 2018; Si et al., 2011; Yan et al., 2021). The integration of machine learning into PHM, especially with the advent of sophisticated algorithms and the increased availability of sensor data opens a new opportunities in RUL prediction. This new technologies is characterized by learning from historical performance data, detecting subtle patterns, and predicting future outcomes with increased accuracy (A. Li et al., 2018; Yang et al., 2020).

The diversity and complexity of data in PHM, combined with the unique operational characteristics of different machinery, pose additional obstacles. These factors complicate the task of creating generalized models that are both accurate and interpretable across varied contexts (Lakkaraju et al., (2016). The need for models that can adapt to such diversity while providing clear insights into their predictions is pressing (Rudin et al., 2022).

In response to these challenges, our research presents a framework that integrates Network Theory with SHapley Additive exPlanations (SHAP) to enhance the accuracy and the interpretability of ML-based RUL predictions. Our contributions are manifold and aim to bridge the gap between complexity and explainability:

I. **Application of a SHAP Theory**: We use a surrogate Model to translate complex interactions from deep learning models into SHAP interaction values.

II. **Application of Network Theory and FIN**: The SHAP interaction values are mapped onto the Feature Interaction Network. This mapping provides a detailed analysis of feature interdependencies to improve our understanding of the factors that affect the reliability of machinery.

III. **Integration of SHAP and Network Theory**: The proposed method combines SHAP values with Network Theory to develop an "augmented" Feature Interaction Network (FIN). This network helps clarify and quantify how different features interact and influence RUL predictions.

The novelty of this research lies in the fusion of network theory with feature importance methodologies to decode the nuanced interplay of operational parameters. By applying a Feature Interaction Network (FIN), a structural map of feature interdependencies further enriched by the integration of SHAP values, we were quantifying and explain feature contributions. Central to this approach was the novel application of surrogate models, facilitating the distillation of SHAP interaction effects into discernible edge strengths within the FIN. Concurrently, the combination of mean absolute SHAP values with network centrality metrics allows positioning a more comprehensive description of feature significance and influence. This research aims to envelope an innovative yet pragmatic set of tools, that can enhance Explainability and interpretability of predictive maintenance practices.

The paper is organized as follows: Section II surveys related literature, establishing the context for our contributions. Section III details our methodology, highlighting the synergistic use of SHAP analysis and Network Theory to decode ML model decisions. Section IV discusses the empirical findings, focusing on the insights gleaned from the FIN and its practical implications for PHM. Section V concludes, reflecting on the impact of our work and suggesting directions for future research in enhancing model transparency and reliability.

2. Background and Literature review

In the field of prognostics and health management (PHM), the ability to accurately predict the Remaining Useful Life (RUL) of machinery is gaining traction (Lei et al., 2018; Ramezani et al., 2019; Zhao & Addepalli, 2020). This increased popularity is largely driven by advancements in machine learning and deep learning technologies. (Berghout & Benbouzid, 2022; Chen, Wu, Zhao, Guretno, Yan, Member, et al., 2021; Ferreira & Gonçalves, 2022). This review aims to summarize recent developments in RUL prediction, highlighting the evolution of methodologies and techniques across various industrial sectors.

The significant improvements in RUL prediction began with the innovative preprocessing of sensor data. For instance, Ensarioğlu et al., (2023) introduced a method that combined difference-based feature construction with a hybrid 1D-CNN-LSTM model, enhancing prediction accuracy significantly. Among the more notable preprocessing techniques is the sliding time window method, which organizes time-series signals into segments of equal length for more consistent input data (Guo et al., 2022). While effective, this method can be labor-intensive and somewhat dependent on the operator's expertise. Another valuable technique is the short-time Fourier transform (STFT), which considers the time correlation of signal sequences, providing a robust basis for subsequent analyses (Liu et al., 2022; Zhang et al., 2023). Also, the integration of long and short-term memory networks (LSTMs) with convolutional block attention modules has improved our understanding of neural decision-making processes (Remadna et al., 2023). The application of deep convolutional variational autoencoders equipped with attention mechanisms has improved the spatial distribution of features, thereby enhancing the interpretability of predictive models (Cheng et al., 2022).

The interpretability of machine learning techniques in RUL prediction has seen significant advancements, particularly through the integration of attention mechanisms and feature fusion frameworks. An attention-based deep learning framework was developed to effectively combine handcrafted and automated features for accurate RUL prediction, demonstrating high efficiency performance on real datasets (Chen, Wu, Zhao, Guretno, Yan, & Li, 2021). Remadna et al., (2023) proposed a fusion of an attention-based convolutional variational autoencoder with an ensemble learning classifier, achieving high accuracy and improved interpretability. Watson (2020) highlighted the conceptual challenges in interpretable machine learning (IML), emphasizing the need for clarity in target definitions and the importance of error rate considerations and testing for IML algorithms. Additionally, Xu et al., (2022) introduced an approaches combined deep learning with other techniques such as particle filters and knowledge distillation to enhance feature extraction, interpretability, and model compression for efficient RUL prediction.

Ye & Yu, (2023) introduced the Selective Adversarial Adaptation Network (SAAN), an approach to domain adaptation employing selective feature interaction for effective knowledge transfer in machine RUL prediction under variable conditions. Kobayashi et al., (2023), also highlighted the critical need for transparency and interpretability in AI models, emphasizing the significance of Explainable AI (XAI) and Interpretable Machine Learning (IML) in RUL prediction in digital twin systems. Zou et al., (2021) proposed an approach for RUL prediction in small data scenarios using a fully convolutional variational auto-encoding network, effectively addressing underfitting issues and demonstrating superior performance in degradation feature extraction and failure threshold determination compared to traditional models.

LIME was proposed by(Ribeiro et al., 2016) as a local model-agnostic approach to interpretability. It has been since then used extensively in prognostics and health management. LIME is a local-model because it approximates the learning model with an interpretable simplified surrogate around a single prediction. As a model0agnostic approach, LIME is a generic and works with any underlying predictive model.

This method has been particularly useful for RUL prediction, as it allows engineers to understand the impact of different features on the predicted outcomes. For instance, Khan et al., (2022); Serradilla Oscar et al., (2020) demonstrated the efficacy of LIME in explaining RUL predictions, enabling a deeper understanding of the degradation patterns and contributing factors, thus facilitating more informed maintenance decisions.

In a recent study, Alomari et al., (2023)developed a comprehensive method for predicting the Remaining Useful Life (RUL) of aircraft engines. Our approach integrates advanced feature engineering, dimensionality reduction through principal component analysis, and a range of feature selection techniques, including Genetic Algorithms, Recursive Feature Elimination, Least Absolute Shrinkage and Selection Operator Regression, and Feature Importances from Random Forest models. A significant innovation in this research is the introduction of the Aggregated Feature Importances with Cross-validation (AFICv) technique. This method enhances the selection process by prioritizing features based on their mean importance also establishes a selection criterion that retains features contributing up to 70% of the cumulative mean sum which is effectively simplifies the model complexity. Another finding in our research is introducing a novel PCA-based interpretability framework to provide actionable insights and enhance the practical utility of our findings for domain experts in the aerospace industry.

2.1. Data Description

The N-CMAPSS dataset (Chao et al., 2021) is a dataset that uses real flight conditions from a commercial jet to simulate the operative conditions (w) within its model. This dataset provides synthetic degradation trajectories for a fleet of turbofan engines, effectively replicating various unknown initial health states under authentic flight conditions. It includes eight distinct datasets derived from 128 engines, each illustrating seven unique failure modes. These modes predominantly affect the flow (F) and efficiency (E) of key engine components such as the fan, low-pressure compressor (LPC), high-pressure compressor (HPC), high-pressure turbine (HPT), and low-pressure turbine (LPT).

Flight conditions within the N-CMAPSS model are categorized into three distinct classes based on the length of the flight. The details of these flight classes, along with the specific failure modes for each dataset, are meticulously documented in Table 1 (4 datasets were used only from the entire original dataset). Additionally, the dataset provides extensive information on the scenario descriptors as in Table 1, and measurements and virtual sensors, which are thoroughly described in the turbofan Jet engine schematic representation Figure 1 and Table 2. This structured approach in modeling the failure modes and operational conditions forms the backbone of the current model development, offering a realistic and detailed perspective of engine degradation under varied flight scenarios.

Table 1 N-CMAPSS Datasets overview

Name	# Units	Flight Classes	Failure Modes
DS01	10	[1 - 2 - 3]	1
DS02	9	[1 - 2 - 3]	2
DS03	15	[1 - 2 - 3]	1
DS05	10	[1 - 2 - 3]	1

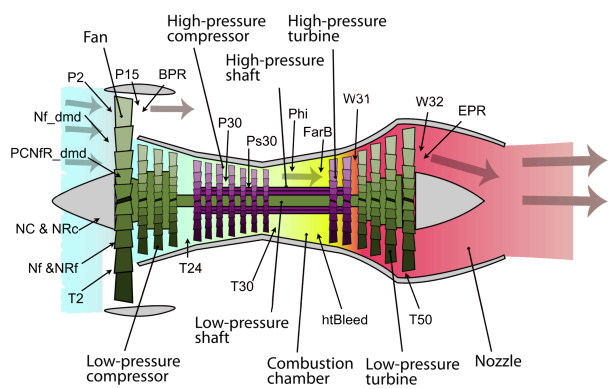

Figure 1 Turbofan Jet Engine Schematic Representation

2.2. Data preprocessing and feature selection

Standardization was applied to the dataset, as detailed in equations 1-3, normalizing each feature to have zero mean and unit variance. This step was essential for ensuring consistency across different data scales and enhancing the efficacy of the subsequent feature selection and machine learning models:

$$\textit{Standardization} \qquad Z = \frac{x - \mu}{\sigma} \tag{1}$$

$$\textit{Mean} \qquad \mu = \frac{1}{n}\sum_{i=1}^{n}(x_i) \tag{2}$$

$$\textit{Standard Deviation} \qquad \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \tag{3}$$

Feature selection is critical in prognostics and health management, offering substantial benefits for applications such as RUL prediction and fault detection. By eliminating redundant features, this process effectively reduces the input dimensions for machine learning models, thereby enhancing their performance by focusing on the most informative attributes (Alomari et al., 2023; Aremu et al., 2020).

In this study, features are selected based on their statistical variability. Sensors that exhibit zero standard deviation, indicating no variation and thus no predictive value, have been excluded. An example of this selection process can be seen in Figure 2, which illustrates how sensors are chosen based on their variability over time. In Figure 2, features such as 'T2', 'W50', and 'Nc' exhibit fluctuating values, whereas other features remain constant, indicating they provide limited informational value to the mode.

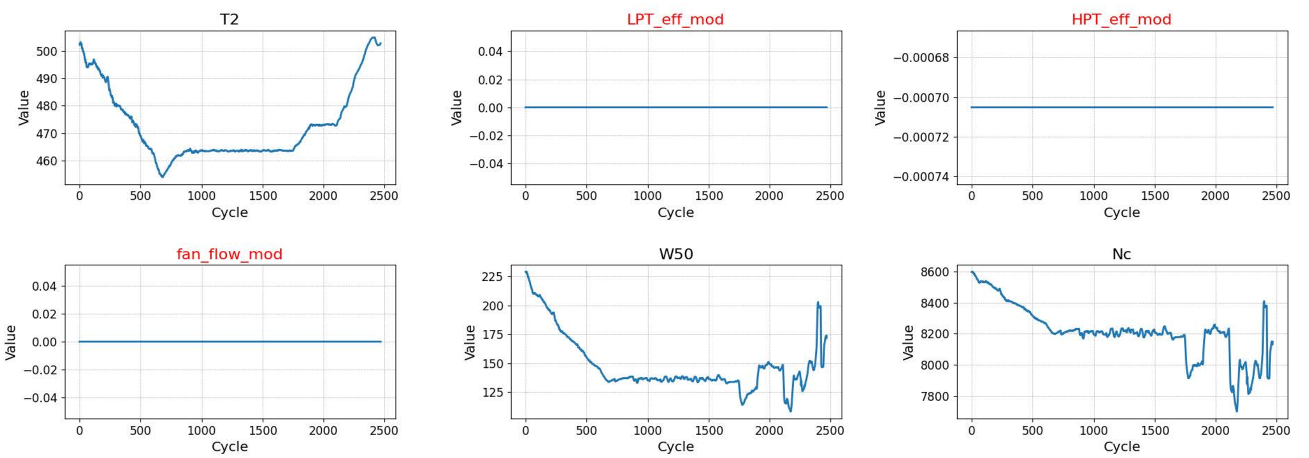

Figure 2 Variability Analysis of Sensor Data for Feature Selection

This selection was specifically tailored to exclude sensors with negligible fluctuations or redundant information, focusing instead on those providing significant insights into engine performance and wear. The final selected features are listed in Table 2.

Table 2 list of the selected features

alt	Mach	TRA	T2	T24
P24	Ps30	P40	P50	Nf
T30	T48	T50	P15	P2
Nc	Wf	T40	P30	P21

3. Methodology

The methodology, illustrated in Figure 3, is based on a composite model integrating Deep Gated Recurrent Units (GRU), Convolutional Neural Networks (CNN), a customized Time Distributed Attention mechanism, and an innovative Feature Interaction Network (FIN). The goals are to improved the precision and interpretability of RUL predictions for aircraft engines.

The GRU layers illustrated in Figure 4 capture the temporal correlations within the sequential engine data, while the CNN layers distill critical spatial features, thereby enhancing the model's capability to identify salient patterns indicative of engine failure. The custom attention layer defined in Figure 5 allows to selectively simplify temporal events within the engine's operational history, further refining the model's predictive accuracy.

To enhance interpretability, the FIN, constructed using SHAP (SHapley Additive exPlanations) values shown in Figure 10, quantifies the impact and interactions of individual features. The node's size within the FIN is representative of the mean absolute SHAP values. This allows better demonstration the feature importance visually. Edge weights are defined by SHAP interaction values, illustrating the strength of the interaction between each pair of features.

The methodological combination of GRU, CNN, attention mechanisms, and SHAP-driven FIN proffers a multidimensional interpretable approach, in the field of aerospace prognostics and health management.

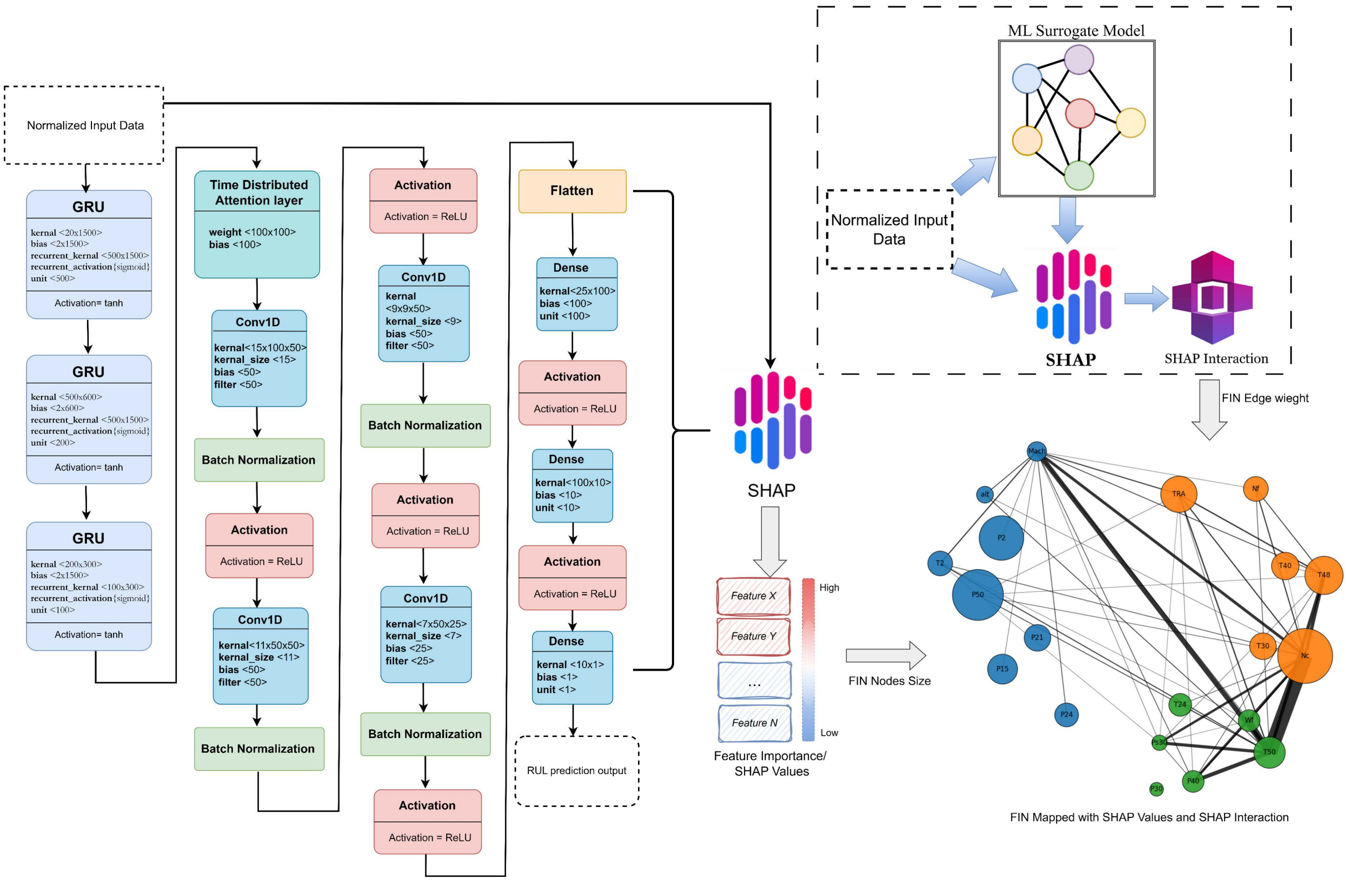

Figure 3 proposed model for RUL prediction and FIN

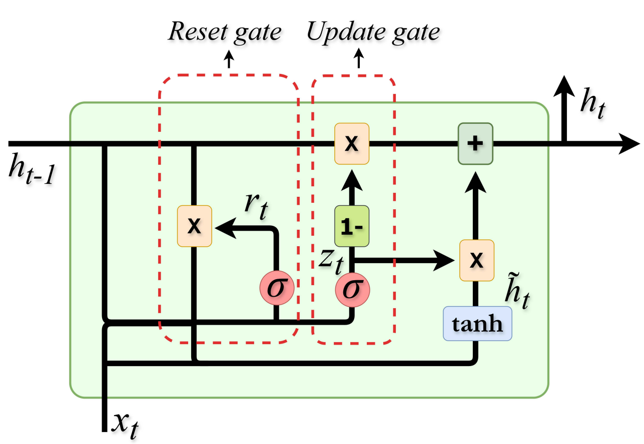

Figure 4 Gated recurrent unit (GRU) neural network structure

The GRU, presented in Figure 4, introduced by (Cho et al., 2014), is a type of recurrent neural network designed to model temporal sequences and long-range dependencies more effectively than standard RNNs. They simplify the recurrent module while retaining the ability to capture dependencies in time-series data, making them computationally efficient and powerful for tasks such as speech recognition, language modeling, and sequential prediction, which are crucial in PHM contexts (Cao et al., 2021; Zhou et al., 2022, Zhou et al., 2023). The core functionality of GRUs relies on the modulation of information flow across sequence steps, controlled by the update and reset gates.

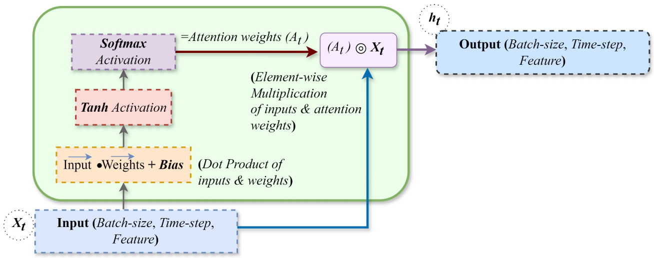

Figure 5 time-distributed attention mechanism

After the GRU layers, a customized Time Distributed Attention mechanism (see Figure 5) was used to improve the model's ability to focus on the most critical features within the sequential data. It applies an attention mechanism to each time step independently. This is achieved by computing an attention score for each feature using a learned weight matrix and bias vector. The scores are then normalized via a softmax function to create attention weights, which are subsequently used to scale the input features. This process allows the model to dynamically prioritize significant information, thereby improving the interpretability and accuracy of RUL predictions.

3.1. SHapley Additive exPlanations (SHAP)

In the field of explainable artificial intelligence (XAI), Shapley Additive exPlanations (SHAP) (Lundberg et al., 2017) values are a central tool for quantifying the contributions of individual features to a model's prediction. Rooted in cooperative game theory, SHAP values, formally described in Equation (4), enable the measurement of each feature's influence by comparing the model's output with and without the presence of the feature. This approach not only fosters transparency but also imbues the analysis with a rigorous mathematical foundation.

SHAP is crucial to PHM (Alomari & Andó, 2024) where understanding the impact of various features on the prediction of system failures or maintenance needs is paramount. SHAP values facilitate this by attributing precise, quantifiable contributions of individual features to the overall prediction of system health, thereby enabling more accurate and timely decision-making.

$$\mathrm{SHAP}(j) = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [f(S \cup \{j\}) - f(S)] \tag{4}$$

S	-	is a subset of features
N	-	is the set of all features
$\lvert S \rvert$	-	denotes the cardinality (size) of set $\boldsymbol{S}$
$f(S \cup \{j\})$	-	is the prediction with both the features in set $\boldsymbol{S}$ and feature $\boldsymbol{j}$
$f(S)$	-	is the prediction with just the features in set $\boldsymbol{S}$

Equation (4) computes the contribution of feature j by iterating over all possible subsets S of the remaining features in N and comparing the difference in the prediction when feature j is included versus when it is excluded.

3.2. Network Theory

Network Theory (Borgatti & Halgin, 2011) provides a framework for understanding the structure and dynamics of complex systems by visualizing them as networks of nodes (features) and edges (interactions). This approach is especially useful in PHM, to reveal the complex interdependencies between system components. By applying Network Theory to create a Feature Interaction Network (FIN), we can perform both visual and quantitative analyses of how individual system features interact and collectively impact overall system behavior (see Figure 6). The decision to use a FIN was deliberate; it helps in mapping out the relationships and dependencies among features effectively and also simplifies the understanding of complex data structures for engineers and domain experts.

To accurately model the interactions within a FIN, the Graphical Lasso (GLasso) algorithm (Friedman et al., 2008) was utilized. GLasso effectively determines the conditional independence structure between variables (features), offering a sparse representation of the feature interaction network. The mathematical formulation of Glasso (Equation 5) is centered on optimizing the following objective function:

$$\min_{\Theta} - \log \det(\Theta) + tr(\mathcal{S}\Theta) + \lambda \|\Theta\|_1 \tag{5}$$

Here, Θ represents the precision matrix (inverse covariance matrix) to be estimated, $\mathcal{S}$ is the empirical covariance matrix of the data, $\log \det(\Theta)$ ensures the positive definiteness of Θ, $tr(\mathcal{S}\Theta)$ is the trace term encouraging fidelity to the observed data, $\|\Theta\|_1$ denotes the L_1 norm imposing sparsity, and λ is a regularization parameter controlling the degree of sparsity. By solving this optimization problem, GLasso identifies significant interactions while discarding the insignificant, resulting in a FIN that highlights the most crucial feature relationships.

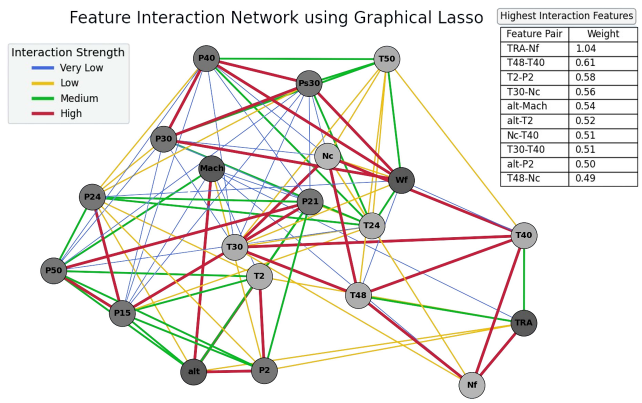

Figure 6 Basic FIN graph with interaction strength

Within the framework of Network Theory, centrality measures serve as tools to compute the prominence of individual nodes. The centrality measures, namely degree centrality for direct linkages, betweenness centrality for intermediary influence, and closeness centrality for overall accessibility are presented in equations (6-8), are instrumental in discerning the structural backbone of the Feature Interaction Network.

Degree centrality $(\boldsymbol{C_D})$ of a node v is defined as the fraction of nodes it is connected to. It reflects the immediate influence of a node within the network.

$$C_D(v) = \frac{\deg(v)}{N-1} \tag{6}$$

Where $\deg(v)$ is the degree of node v (i.e., the number of edges incident to v and N is the total number of nodes in the network. Degree centrality helps us pinpoint features that exert considerable control over the system's behavior, thereby identifying potential points of proactive maintenance and intervention.

Betweenness centrality $(\boldsymbol{C_B})$ of a node v quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

$$C_B(v) = \sum_{\mathcal{S} \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{7}$$

Where σ_{st} is the total number of shortest paths from node $\mathcal{S}$ to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v, it highlights the node's role in facilitating interactions between other nodes. Identifying such nodes helps in strategizing interventions that can prevent cascading failures in engine operations.

Closeness centrality $\boldsymbol{C_c}(v)$ measures how close a node is to all other nodes in the network, indicating how easily information can flow from the given node to others.

$$C_c(v) = \frac{N-1}{\sum_{u \neq v} d(v,u)} \tag{8}$$

Where $d(v,u)$ is the shortest path distance between nodes v and u, and N is the total number of nodes in the network. Features with high closeness centrality are likely to affect the system more rapidly, making them critical targets for monitoring and early preventive maintenance.

Community detection algorithms, such as the Louvain method (De Meo et al., 2011) given by equation (9), partition the network into communities or clusters of nodes that are more densely connected with each other than with the rest of the network. This segmentation can reveal modular structures within the feature set, suggesting subsystems within the engine that have distinct behaviors.

$$Q = \frac{1}{2M} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{9}$$

The following definitions characterize the parameters of the algorithms

- A_{ij} represents the weight of the edge between nodes i and j. For unweighted networks, A_{ij} is 1 if there is an edge between i and j, and 0 otherwise.
- k_i *and* k_j are the sum of the weights of the edges attached to nodes i and j, respectively.
- m is the sum of all the edge weights in the network.
- c_i *and* c_j are the communities of nodes i and j
- δ is the Kronecker delta function, which is 1 if $c_i = c_j$ (i.e., nodes i and j are in the same community) and 0 otherwise.

The goal of the Louvain method is to maximize Q through a heuristic approach that iteratively groups nodes into communities.

3.3. Evaluation metrics

The proposed model evaluation was conducted using the N-CMAPSS datasets (DS01, DS02, DS03 and DS05), focusing on the accuracy of Remaining Useful Life (RUL) predictions. The performance of our proposed model was primarily assessed by measuring the discrepancy between the predicted and actual RUL values. For this purpose, we employed two key metrics: Root Mean Square Error (RMSE) and NASA Score (Saxena et al., 2008), as defined in equations (10 - 13). These metrics, calculated over the number of data points (n), provided a comprehensive understanding of the model's predictive accuracy and reliability in various operational scenarios presented within the N-CMAPSS datasets.

$$RMSE\ (P_{RUL}, T_{RUL}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{RUL} - T_{RUL})^2} \tag{10}$$

$$\text{NASA Score}_i = \begin{cases} exp\left(-\frac{\Delta_i}{10}\right) - if\ \Delta_i < 0 \\ exp\left(-\frac{\Delta_i}{13}\right) - if\ \Delta_i \geq 0 \end{cases} \tag{11}$$

Where:

$$\Delta_i = RUL_{predicted\ i} - RUL_{true_i} \tag{12}$$

$$NASA\ Score = \frac{1}{n}\sum_{i=1}^{n} NASA\ Score_i \tag{13}$$

4. Results and Discussion

Our enhanced approach to predicting Remaining Useful Life (RUL) uses an integration of Deep Gated Recurrent Units (GRU), Convolutional Neural Networks (CNN), and a custom Time Distributed Attention mechanism. This tailored combination has advanced the accuracy of RUL predictions by effectively capturing complex temporal and spatial patterns within engine operational data, crucial for early and accurate fault detection. The inclusion of the Custom Attention Layer allows for identifying critical features and time steps, significantly refining the interpretability of our predictive models. The effectiveness of these innovations is substantiated by our empirical results presented in Tables 3 and 4. Table 3 includes a comparison of some of our results with three methods from the literature, while Table 4 presents the results for the remaining engines, for which no direct comparisons to existing studies could be made.

Across the different datasets, the model demonstrates proficiency in RUL prediction, as evidenced by the calculated Root Mean Square Error (RMSE) and the NASA prognostics score, with a significant performance in the critical RUL phase. This is important since the latter half of life where accurate prediction is most vital. Particularly significant are the outcomes on DS02 and DS03, where the model achieves RMSE values as low as 2 cycles for the critical RUL, alongside correspondingly low NASA-scores, highlighting the model's precision in the most consequential phase of the engine's lifecycle.

The visualization of the RUL prediction and critical RUL of two engines, 9 and 12, from DS01 and DS03, respectively, along with their SHAP values, is presented in Figures 7 and 8. These figures illustrate the model's ability to track the Remaining Useful Life (RUL) over engine cycles accurately, with a particular focus on the critical RUL phase. The SHAP interpretation plots highlight the influence of various sensors on the model's predictions.

For Engine 12, significant features include 'Nc' (corrective speed), 'P50' (pressure at the fan outlet), and 'T2' (temperature at the fan inlet). The high SHAP values for these features indicate their substantial impact on the RUL predictions. Specifically, 'Nc' demonstrates a strong correlation with the engine's operational efficiency, reflecting its role in adaptive speed control. Similarly, 'P50' and 'T2' provide crucial insights into the pressure and temperature dynamics, essential for accurate prognostics.

In Engine 9, the SHAP values reveal 'Nc', 'T50' (temperature at the engine outlet), and 'P2' (pressure at the fan inlet) as key contributors. The interactions between 'Nc' and 'T50' (as they have opposite influence) suggest that the corrective speed adjustments are heavily influenced by thermal conditions at critical engine points. The significant SHAP values for 'P2' underscore the importance of pressure measurements in anticipating engine failures.

Table 3 Prognostics performance assessment comparison with different methods

Dataset	**Engine**	RMSE Proposed	RMSE Literature (Koutroulis et al., 2022)	RMSE Literature LR+ (Maulana et al., 2023)	RMSE Literature MLP+ (Maulana et al., 2023)
DS02	11	4.7	5.1	11.4	11.5
	14	6.1	11.9	10.9	11.1
	15	4	5.8	8.9	18.2
DS03	13	3.9	6.8	--	--
	14	3.2	5.1	--	--
	15	2.1	3.04	--	--

Table 4 Comparative assessment of RMSE and NASA-Score metrics for RUL prediction across engine units

Dataset	Engine	RMSE	RMSE Critical RUL	NASA-Score	NASA-Score Critical RUL
DS01	**7**	8.4	7	1.1	0.9
	8	6	4	0.6	0.5
	9	14	12	2.1	1.5
	10	5	3.5	0.5	0.37
DS03	**10**	8	2.9	0.8	0.22
	11	8	3	0.8	0.25
	12	17	7.3	6.4	1
DS05	**7**	10	3.8	1.9	0.37
	8	6	2.4	0.7	0.2
	9	7	3.6	0.8	0.27
	10	9	3.8	1.2	0.28

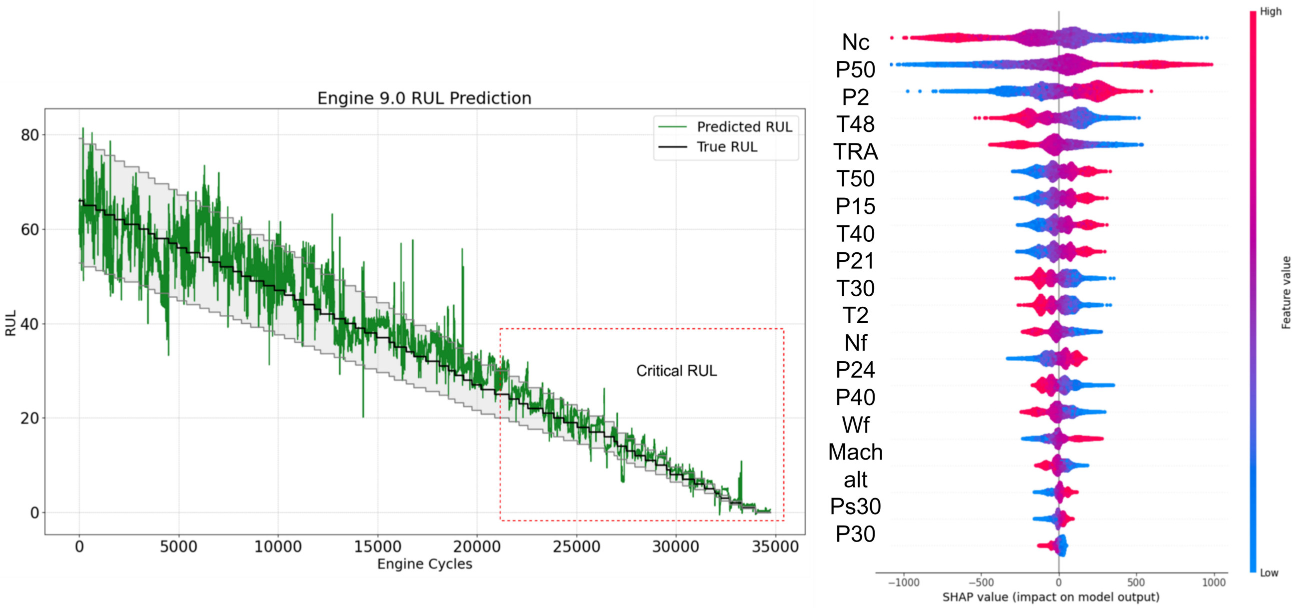

Figure 3 RUL prediction of engine 9 of DS01 with SHAP summary

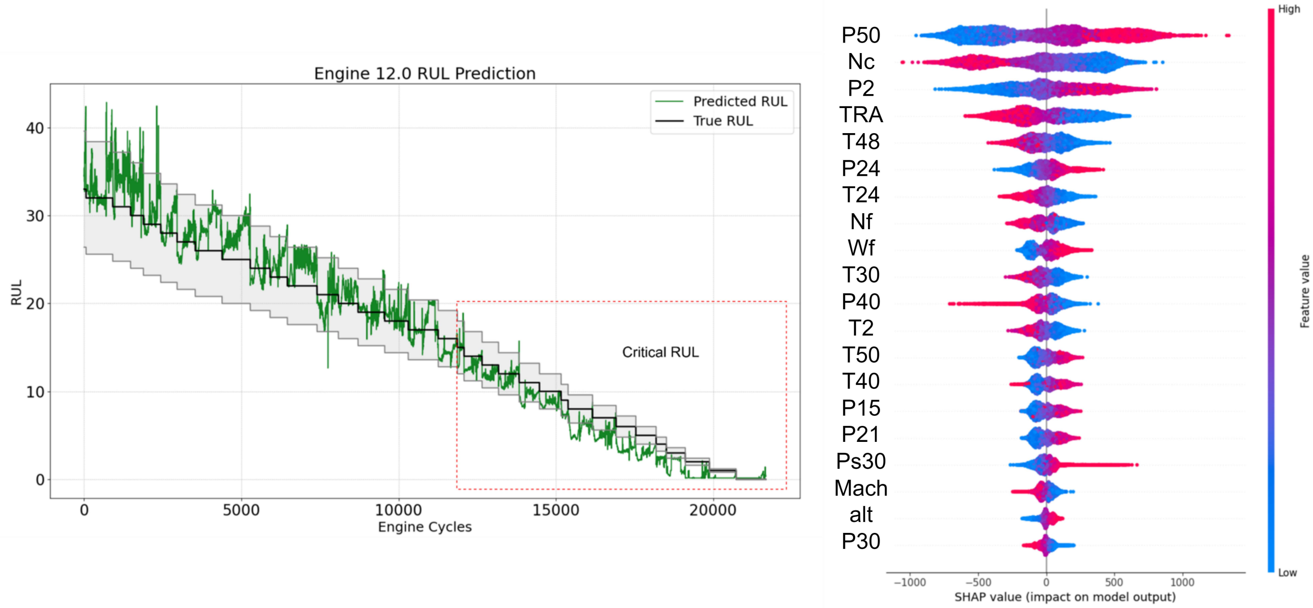

Figure 4 RUL prediction of engine 12 of DS03 with SHAP summary

The Feature Interaction Network (FIN) in Figure 8 provides an overview of the complex relationships inherent in the proposed predictive model for Remaining Useful Life (RUL). Through community detection algorithms, it has discerned distinct clusters within the network, indicative of underlying structures where subsets of features exhibit tightly knit interactions, potentially alluding to functional modules within the engine's operational parameters. The community color-coding allows to observe the modular nature of feature interdependencies, which may correspond to different physical or operational aspects of engine performance. Additionally, the betweenness centrality analysis reveals key nodes such as 'TRA,' 'P24,' and 'P15' that act as critical conduits in the flow of information through the network, signifying their roles in the model's inference processes.

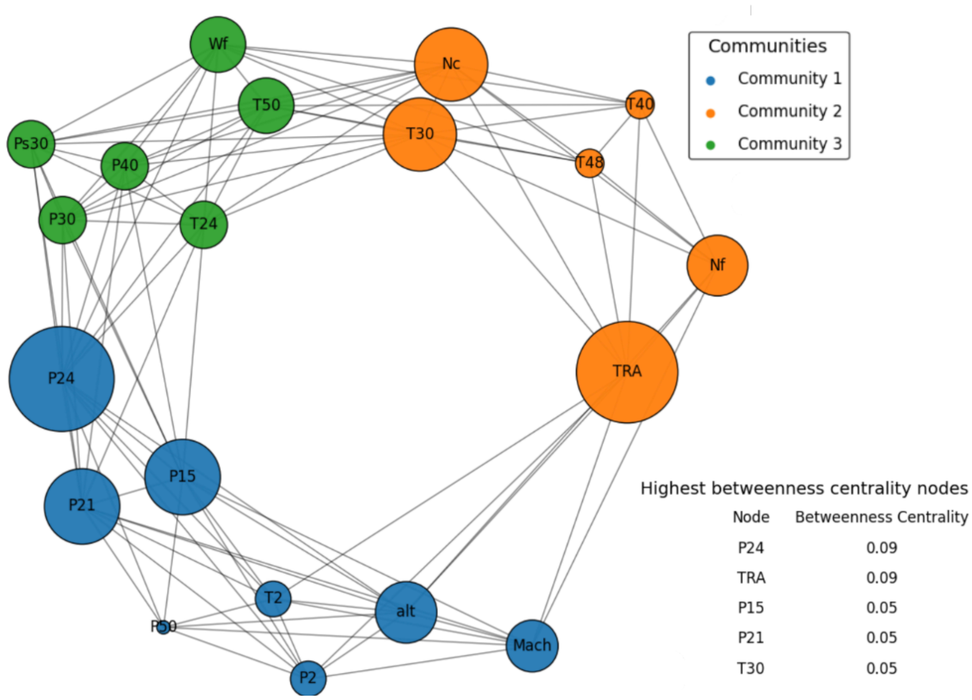

Figure 5 FIN with betweenness centrality and community detection

In Figure 10, we illustrate an innovative Feature Interaction Network (FIN) that leverages SHAP values to illuminate the complex dynamics within our predictive model for Remaining Useful Life (RUL) of aerospace engines. This network diagram, empowered by a surrogate model, not only visualizes the relative influence of various engine features but also clarifies their interrelationships. Each node, scaled according to mean absolute SHAP values, reflects the magnitude of influence each feature holds over the RUL predictions, with larger nodes marking more influential features.

These nodes are distinctly color-coded to represent different communities or clusters of features that share similar behavior patterns within the predictive framework, highlighting how groups of related features collectively impact engine performance. The edges between nodes, whose thickness is determined by the SHAP interaction values, illustrate the strength of the interactions between feature pairs, revealing critical dependencies and synergies.

Key interactions such as those between 'NC' (corrective speed) and temperatures at critical engine locations ('T50' and 'T48') suggest a profound connection between engine speed adjustments and thermal conditions. This relationship is crucial for maintaining optimal engine performance, particularly under varying operational stresses. The interaction between 'NC' and 'T50' highlights how adjustments in engine speed can be crucial in managing the engine's thermal output to avoid overheating while maintaining efficiency.

Further, the interaction between 'T50' and 'Mach' (aircraft velocity relative to the speed of sound) underscores the significant impact of aerodynamic performance on engine thermal management. The relationship between engine thermal outputs and flight speed suggests that higher speeds may require adjustments in thermal management strategies to maintain engine integrity and performance.

Additionally, the 'NC - Mach' interaction points to a dynamic balancing act required between engine speed and aircraft velocity, indicating that engine control systems need to be highly adaptive to changes in flight dynamics. This adaptiveness is crucial for optimizing fuel consumption and minimizing wear and tear under different flight conditions.

Lastly, the interaction between 'T50' and 'P40' (pressure at the fan outlet) sheds light on how temperature and pressure management are interlinked, playing a pivotal role in ensuring the engine's thrust efficiency and overall stability. This insight is particularly valuable for developing more effective predictive maintenance strategies, aiming to reduce unexpected downtimes and extend the engine's useful life.

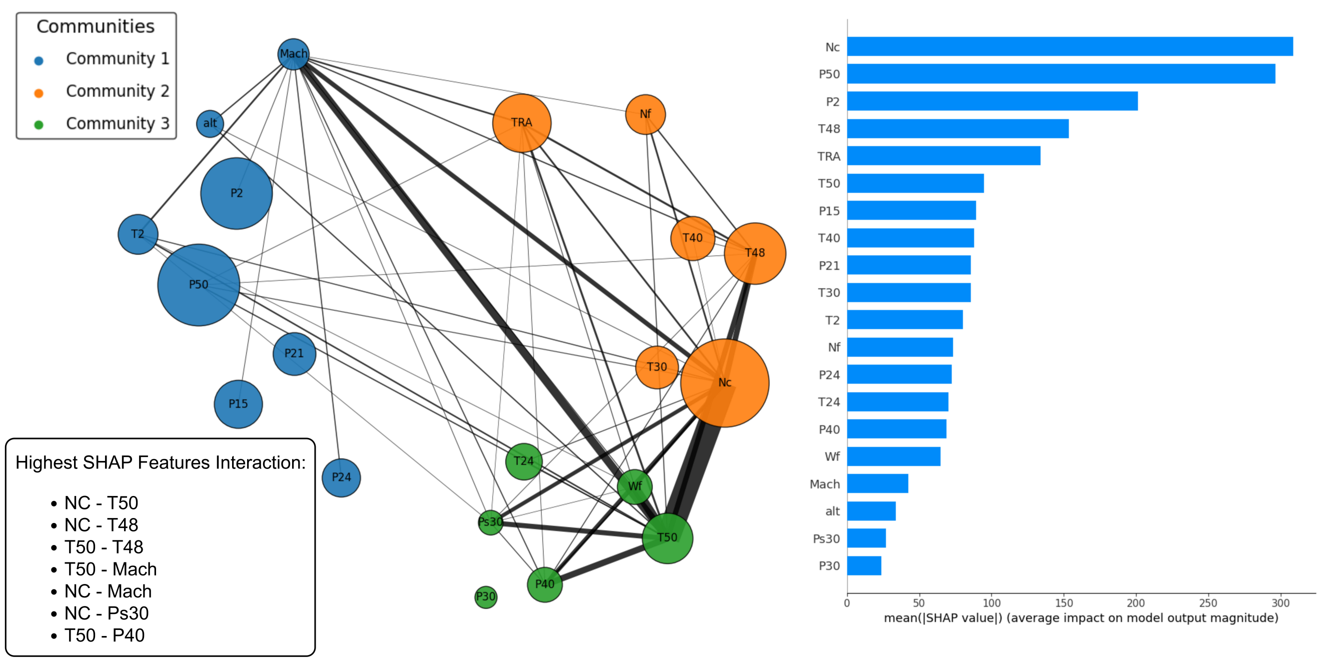

Figure 6 Feature Interaction Network (FIN) Visualizing Key Dependencies and Community Structures: This network map illustrates the Feature Interaction Network (FIN) with nodes sized according to their mean absolute SHAP values, which reflect the impact magnitude on the model's output. The nodes are color-coded by community, identifying clusters of tightly interconnected features that influence system behavior in distinct ways. Thicker lines between nodes indicate stronger SHAP

feature interactions, highlighting critical dependencies such as NC-T50 and T50-Mach, which are pivotal for understanding complex dynamics within the aerospace engine's operations

5. CONCLUSION

In conclusion, our research aims to advance the predictive maintenance field by developing a prognostic framework that combines cutting-edge machine learning techniques with innovative interpretative methodologies to predict the Remaining Useful Life (RUL) of aerospace engines. Utilizing a Surrogate Model, we have successfully mapped complex SHAP feature interactions into a well-defined Feature Interaction Network (FIN). This network, structured with nodes proportionally scaled by mean absolute SHAP values and connections defined by the strength of SHAP interactions, vividly represents the intricate relationships between operational parameters.

Our detailed analysis highlighted crucial feature interactions, notably between corrective speed and critical engine temperature, which are point factors essential for optimizing engine efficiency and performance. Furthermore, the application of community detection in the FIN has significantly deepened our understanding of these features, grouping related variables to illuminate how they collectively impact RUL predictions. This clustering clarifies the predictive model's structure and enhances the interpretability of the data, providing clear pathways for intervention.

The visual representation of the FIN is not merely an analytical tool; it acts as a vital conduit translating complex, data-driven insights into tangible, operational strategies. This visualization underscores the transformative potential of interpretative machine learning to convert abstract data into actionable intelligence, a resource of value in the high-stakes field of aerospace prognostics where the accuracy of predictions can directly influence operational safety and maintenance efficiency.

ACKNOWLEDGMENT

Project no. TKP2021-NVA-29 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme

REFERENCES

Alomari, Y., & Andó, M. (2024). SHAP-based insights for aerospace PHM: Temporal feature importance, dependencies, robustness, and interaction analysis. *Results in Engineering*, *21*. https://doi.org/10.1016/j.rineng.2024.101834

Alomari, Y., Andó, M., & Baptista, M. L. (2023a). Advancing aircraft engine RUL predictions: an interpretable integrated approach of feature engineering and aggregated feature importance. *Scientific Reports 2023 13:1*, *13*(1), 1–14. https://doi.org/10.1038/s41598-023-40315-1

Alomari, Y., Andó, M., & Baptista, M. L. (2023b). Advancing aircraft engine RUL predictions: an interpretable integrated approach of feature engineering and aggregated feature importance. *Scientific Reports 2023 13:1*, *13*(1), 1–14. https://doi.org/10.1038/s41598-023-40315-1

Aremu, O. O., Cody, R. A., Hyland-Wood, D., & McAree, P. R. (2020). A relative entropy based feature selection framework for asset data in predictive maintenance. *Computers & Industrial Engineering*, *145*, 106536. https://doi.org/10.1016/J.CIE.2020.106536

Baptista, M. L., Goebel, K., & Henriques, E. M. P. (2022). Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. *Artificial Intelligence*, *306*, 103667. https://doi.org/10.1016/J.ARTINT.2022.103667

Berghout, T., & Benbouzid, M. (2022). A Systematic Guide for Predicting Remaining Useful Life with Machine Learning. *Electronics 2022*, *Vol. 11*, *Page 1125*, *11*(7), 1125. https://doi.org/10.3390/ELECTRONICS11071125

Borgatti, S. P., & Halgin, D. S. (2011). On Network Theory. *Organization Science*, *22*(5), 1168–1181. https://doi.org/10.1287/orsc.1100.0641

Calabrese, F., Regattieri, A., Botti, L., Mora, C., & Galizia, F. G. (2020). Unsupervised fault detection and prediction of remaining useful life for online prognostic health management of mechanical systems. *Applied Sciences (Switzerland)*, *10*(12), 4120. https://doi.org/10.3390/APP10124120

Cao, Y., Jia, M., Ding, P., & Ding, Y. (2021). Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network. *Measurement*, *178*, 109287. https://doi.org/10.1016/J.MEASUREMENT.2021.109287

Chao, M., Kulkarni, C., Goebel, K., & Fink, O. (2021). Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics. *NASA Ames Research Center, Moffett Field*.

Chen, Z., Wu, M., Zhao, R., Guretno, F., Yan, R., & Li, X. (2021). Machine Remaining Useful Life Prediction via an Attention-Based Deep Learning Approach. *IEEE Transactions on Industrial Electronics*, *68*(3), 2521–2531. https://doi.org/10.1109/TIE.2020.2972443

Chen, Z., Wu, M., Zhao, R., Guretno, F., Yan, R., Member, S., & Li, X. (2021). Machine Remaining Useful Life

Prediction via an Attention-Based Deep Learning Approach. *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, *68*(3). https://doi.org/10.1109/TIE.2020.2972443

Cheng, Y., Hu, K., Wu, J., Zhu, H., & Shao, X. (2022). Autoencoder Quasi-Recurrent Neural Networks for Remaining Useful Life Prediction of Engineering Systems. *IEEE/ASME Transactions on Mechatronics*, *27*(2), 1081–1092. https://doi.org/10.1109/TMECH.2021.3079729

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *ArXiv Preprint ArXiv*.

De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2011). *Generalized Louvain Method for Community Detection in Large Networks*. https://doi.org/10.1109/ISDA.2011.6121636

Deutsch, J., & He, D. (2018). Using deep learning-based approach to predict remaining useful life of rotating components. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *48*(1), 11–20. https://doi.org/10.1109/TSMC.2017.2697842

Duc Nguyen, V., Kefalas, M., Yang, K., Apostolidis, A., Olhofer, M., Limmer, S., & Bäck, T. (2019). A Review: Prognostics and Health Management in Automotive and Aerospace. *International Journal of Prognostics and Health Management*, *10*(2), 35. https://www.klm.com/corporate/en/publications/2015

Ensarioğlu, K., İnkaya, T., & Emel, E. (2023). Remaining Useful Life Estimation of Turbofan Engines with Deep Learning Using Change-Point Detection Based Labeling and Feature Engineering. *Applied Sciences 2023, Vol. 13, Page 11893*, *13*(21), 11893. https://doi.org/10.3390/APP132111893

Ferreira, C., & Gonçalves, G. (2022). Remaining Useful Life prediction and challenges: A literature review on the use of Machine Learning Methods. *Journal of Manufacturing Systems*, *63*, 550–562. https://doi.org/10.1016/J.JMSY.2022.05.010

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Guo, R., Li, H., Huang, C., Zhao, C., Huang, X., Liu, H., Li, J., & Chen, R. (2022). A sequence-to-sequence remaining useful life prediction method combining unsupervised LSTM encoding-decoding and temporal convolutional network. *Measurement Science and Technology*, *33*(8), 085013. https://doi.org/10.1088/1361-6501/AC632D

Khan, T., Ahmad, K., Khan, J., Khan, I., & Ahmad, N. (2022). An Explainable Regression Framework for Predicting Remaining Useful Life of Machines. *2022 27th International Conference on Automation and Computing: Smart Systems and Manufacturing, ICAC 2022*. https://doi.org/10.1109/ICAC55051.2022.9911162

Kobayashi, K., Almutairi, B., Sakib, M. N., Chakraborty, S., & Alam, S. B. (2023). *Explainable, Interpretable & Trustworthy AI for Intelligent Digital Twin: Case Study on Remaining Useful Life*. https://arxiv.org/abs/2301.06676v1

Kononov, E., Klyuev, A., & Tashkinov, M. (2023). Prediction of Technical State of Mechanical Systems Based on Interpretive Neural Network Model. *Sensors 2023, Vol. 23, Page 1892*, *23*(4), 1892. https://doi.org/10.3390/S23041892

Koutroulis, G., Mutlu, B., & Kern, R. (2022). Constructing robust health indicators from complex engineered systems via anticausal learning. *Engineering Applications of Artificial Intelligence*, *113*. https://doi.org/10.1016/j.engappai.2022.104926

Lakkaraju, H., Bach, S. H., & Leskovec, J. (2016). Interpretable Decision Sets: A Joint Framework for Description and Prediction. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1675–1684. https://doi.org/10.1145/2939672.2939874

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

Li, A., Yang, X., Dong, H., Xie, Z., & Yang, C. (2018). Machine learning-based sensor data modeling methods for power transformer PHM. *Sensors (Switzerland)*, *18*(12). https://doi.org/10.3390/s18124430

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

Liu, B., Gao, Z., Lu, B., Dong, H., & An, Z. (2022). *SAL-CNN: Estimate the Remaining Useful Life of Bearings Using Time-frequency Information*. https://arxiv.org/abs/2204.05045v1

Lundberg, S. M., Allen, P. G., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*. https://github.com/slundberg/shap

Maulana, F., Starr, A., & Ompusunggu, A. P. (2023). Explainable Data-Driven Method Combined with Bayesian Filtering for Remaining Useful Lifetime Prediction of Aircraft Engines Using NASA CMAPSS Datasets. *Machines*, *11*(2). https://doi.org/10.3390/machines11020163

Ramezani, S., Moini, A., & Riahi, M. (2019). Prognostics and Health Management in Machinery: A Review of Methodologies for RUL prediction and Roadmap. *International Journal of Industrial Engineering &*

Supply Chain Management, *6*(1), 38–61. www.ijiems.com

Remadna, I., Terrissa, L. S., Al Masry, Z., & Zerhouni, N. (2023). RUL Prediction Using a Fusion of Attention-Based Convolutional Variational AutoEncoder and Ensemble Learning Classifier. *IEEE Transactions on Reliability*, *72*(1), 106–124. https://doi.org/10.1109/TR.2022.3190639

Ren, C., Li, H., Zhang, Z., & Si, X. (2023). A Remaining Useful Life Prediction Method with Degradation Model Calibration. *Proceedings of 2023 IEEE 12th Data Driven Control and Learning Systems Conference, DDCLS 2023*, 172–177. https://doi.org/10.1109/DDCLS58216.2023.10166929

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2939672.2939778

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistics Surveys*, *16*, 1–85. https://doi.org/10.1214/21-SS133

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. *2008 International Conference on Prognostics and Health Management, PHM 2008*. https://doi.org/10.1109/PHM.2008.4711414

Serradilla Oscar, Ekhi Zugasti, Carlos Cernuda, Andoitz Aranburu, Julian Ramirez de Okariz, & Urko Zurutuza. (2020). Interpreting Remaining Useful Life Estimations Combining Explainable Artificial Intelligence and domain knowledge in industrial machinery. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–8.

Si, X. S., Wang, W., Hu, C. H., & Zhou, D. H. (2011). Remaining useful life estimation – A review on the statistical data driven approaches. *European Journal of Operational Research*, *213*(1), 1–14. https://doi.org/10.1016/J.EJOR.2010.11.018

Vollert, S., Atzmueller, M., & Theissler, A. (2021). Interpretable Machine Learning: A brief survey from the predictive maintenance perspective. *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA, 2021-September*. https://doi.org/10.1109/ETFA45728.2021.9613467

Watson, D. (2020). Conceptual challenges for interpretable machine learning. *Synthese*, *200*, 1–33. https://doi.org/10.2139/SSRN.3668444

Xu, Q., Chen, Z., Wu, K., Wang, C., Wu, M., & Li, X. (2022). KDnet-RUL: A Knowledge Distillation Framework to Compress Deep Neural Networks for Machine Remaining Useful Life Prediction. *IEEE Transactions on Industrial Electronics*, *69*(2), 2022–2032. https://doi.org/10.1109/TIE.2021.3057030

Yan, B., Ma, X., Huang, G., & Zhao, Y. (2021). Two-stage physics-based Wiener process models for online RUL prediction in field vibration data. *Mechanical Systems and Signal Processing*, *152*. https://doi.org/10.1016/j.ymssp.2020.107378

Yang, F., Zhang, W., Tao, L., & Ma, J. (2020). Transfer learning strategies for deep learning-based PHM algorithms. *Applied Sciences (Switzerland)*, *10*(7). https://doi.org/10.3390/app10072361

Ye, Z., & Yu, J. (2023). A Selective Adversarial Adaptation Network for Remaining Useful Life Prediction of Machines Under Different Working Conditions. *IEEE Systems Journal*, *17*(1), 62–71. https://doi.org/10.1109/JSYST.2022.3183134

Zhang, K. ;, Liu, R., Zhang, K., & Liu, R. (2023). LSTM-Based Multi-Task Method for Remaining Useful Life Prediction under Corrupted Sensor Data. *Machines 2023, Vol. 11, Page 341*, *11*(3), 341. https://doi.org/10.3390/MACHINES11030341

Zhao, Y., & Addepalli, S. (2020). *ScienceDirect ScienceDirect Remaining Useful Life Prediction using Deep Learning Approaches: A Review*. https://doi.org/10.1016/j.promfg.2020.06.015

Zhou, J., Qin, Y., Chen, D., Liu, F., & Qian, Q. (2022). Remaining useful life prediction of bearings by a new reinforced memory GRU network. *Advanced Engineering Informatics*, *53*, 101682. https://doi.org/10.1016/J.AEI.2022.101682

Zhou, J., Qin, Y., Luo, J., & Zhu, T. (2023). Remaining Useful Life Prediction by Distribution Contact Ratio Health Indicator and Consolidated Memory GRU. *IEEE Transactions on Industrial Informatics*, *19*(7), 8472–8483. https://doi.org/10.1109/TII.2022.3218665

Zou, Y., Shi, K., Liu, Y., Zhangjidong, Liu, Y., & Ding, G. (2021). A Novel Machine RUL Prediction Method in Small Sample based on Fully Convolutional Variational Auto-Encoding Network. *2021 26th International Conference on Automation and Computing: System Intelligence through Automation and Computing, ICAC 2021*. https://doi.org/10.23919/ICAC50006.2021.9594066

BIOGRAPHIES

Yazan Alomari is a Ph.D. candidate in Computer Science at Eötvös Loránd University, Hungary, since September 2020, with collaboration from TU Delft University, The Netherlands. His research focuses on leveraging large language models for Prognostics and Health Management (PHM) in aerospace, emphasizing predictive maintenance through AI insights. His work spans across artificial intelligence, machine learning, and data mining & analysis, aiming to enhance system reliability and operational efficiency. Yazan has also contributed to academia as a University Lecturer at Eötvös Loránd University, imparting knowledge in data science, machine learning, and digital manufacturing. His publications underscore advancements in aerospace PHM, eXplainable AI and machine learning interpretation, showcasing his commitment to advancing AI applications in critical sectors.

Marcia L. Baptista BSc. and MSc. in Informatics and Computer Engineering. Instituto Superior Tecnico, Lisbon, Portugal September 2008) is an Assistant Professor at the Aerospace Faculty of TU Delft since 2020. She holds a PhD from the Engineering Design and Advanced Manufacturing (EDAM) program under the umbrella of MIT Portugal. Her research focuses on the development of prognostics techniques for aeronautics equipment. Her research interests include eXplainable Artificial Intelligence (xAI), machine learning, hybrid modeling, maintenance and prognostics.

Mátyás Andó habil., Ph.D., is Associate Professor in Institute of Computer Science, at Faculty of Informatics of Eötvös Loránd University, ELTE, Budapest, Hungary. He teaches: Manufacturing Technologies, CNC programing, CAM systems. His research area: Improve the efficiency in the industry, I4.0 solutions (CNC machines, PLC, Welding, Image processing). Material sciences (polymer development, tribology, 3D printing). Membership: Hungarian Association of Mechanical Engineers – GTE, members of the Hungarian Academy of Sciences (Section of Engineering Sciences).