Linking Thermodynamic Modelling and Explainable AI: A Framework for Long-Horizon Life Prediction

Ahmad Kamal Bin Mohd Nor * and Masdi Muhammad

Posted Date: 29 April 2026

doi: 10.20944/preprints202604.2033.v1

Keywords: explainable AI; AI interpretability; shapley values; prognostics; gas turbine; long term forecasting

 doi:10.20944/preprints202604.2033.v1

Article

Linking Thermodynamic Modelling and Explainable AI: A Framework for Long-Horizon Life Prediction

Ahmad Kamal bin Mohd Nor * and Masdi Muhammad

Mechanical Engineering Department, Universiti Teknologi Petronas, Perak, Malaysia

* Correspondence: kamal.mohd@utp.edu.my

Abstract

Error-proof prediction is currently a major interest in machine learning (ML) based-gas turbine (GT) failure prognostics applications, indicated by the rise of probabilistic, ensemble, Physics-informed, and explainable AI (XAI) models. For effective maintenance planning, it is important to validate the existence of degradation during life assessment. However, probabilistic and ensemble models can only confirm anomaly which does not necessarily point to degradation while Physics-informed models sometimes work poorly on actual data due to limitations of physics models. XAI can make ML model transparent to confirm the presence of degradation. Existing XAI-based GT prognostics works however suffer from the lack of uncertainty quantification, making it hard to evaluate the prediction trustworthiness. Subsequently, false explanation, which misguides maintenance decision making, risked being generated. In this work, a transparent machine learning (ML) model that predicts and justifies gas turbine's remaining useful life (RUL) prediction is developed, evaluated and validated using fouling failure created from thermodynamic modelling. Specifically, a Bayesian ML model incorporated with XAI capability was employed to estimate the RUL of a twinshaft GT. Thermodynamic modelling was conducted on actual GT data and compressor fouling was injected to create failure data. The uncertainty and trend from the ML prediction and the generated XAI explanation were compared with baseline uncertainty level and explanation to confirm anomaly occurrence to support RUL prediction. The life estimation and explanation were used next to determine the defective component. The model predicted MAPE metric to be 18.04% in a multi-step ahead, long term forecasting horizon. The predictions are supported by the uncertainty level of 0.146 and 0.147 for partial and failure data respectively which is higher than the baseline level of 0.022 that implies anomaly. The prediction and explanations match the thermodynamic modelling which points to compressor failure.

Keywords: explainable AI; AI interpretability; shapley values; prognostics; gas turbine; long term forecasting

I. Introduction

Gas turbines (GT) are used in electricity energy generation, aircraft thrust production and naval propulsion [1,2]. Industrial GT is generally composed of gas compressor to compress intake air, combustor where high pressure mixture of air and fuel is heated and a turbine to drive the compressor. In a twin-shaft GT with a free power turbine configuration, a free power turbine is added to the GT.

In present day, data driven methods such as machine learning (ML) in general, and more specifically deep learning (DL) models, play an increasing part in GT failure prognostics research due to limitations of physics-based models [3]. Traditional data driven methods, however, could give false RUL predictions due to changes in testing data characteristics, data contamination, sensors malfunction, or model development mistakes, amongst others, that could lead to erroneous forecasts. Unlike physics-based models, it is impossible to know the root cause of the prediction error, as ML are black-box methods.

 doi:10.20944/preprints202604.2033.v1

Error-proof estimation has thus become the major challenge in data-driven-based GT failure prognostic. This is shown by the recent rise in probabilistic [4], ensemble [5], Physics-informed [6] and explainable AI (XAI)-based approaches [7] which provide better reasoning in degradation estimation. Nevertheless, there are some obvious weaknesses in these models that need urgent attention:

- Probabilistic models can only indicate anomaly especially when the model is trained uniquely using healthy data, due to absence of failure data [8,9]. Physics-informed method is much harder to be effective with realistic data due to oversimplification and assumptions of physical phenomena [10,11]. Meanwhile, ensemble models average prediction measure by integrating several models to obtain a stronger model. Thus, they also act as probabilistic models [12]. In other words, these models lack of transparency make it difficult for any maintenance decision making.
- XAI, on the other hand, is able to make machine learning black-box model transparent. XAI approaches, however, risk giving false explanation if the developed model does not represent well the data. This can happen if the model is not well trained or trained with insufficient data.

In this work, a transparent machine learning (ML) model that predicts and justifies GT life prediction is developed, evaluated and validated using fouling failure created from thermodynamic modelling. Specifically, a Bayesian ML model incorporated with XAI capability was employed to estimate the RUL of a twinshaft GT. Thermodynamic modelling was conducted on actual GT data and compressor fouling was injected to create failure data. The uncertainty and trend from the ML prediction and the generated XAI explanation were compared with baseline uncertainty level and explanation to confirm anomaly occurrence to support RUL prediction. The life estimation and explanation were used next to determine the defective component.

The main objectives of this research are as thus follows:

1. To develop a transparent ML model that predicts and explains GT multistep ahead, continuous RUL prediction.
2. To validate the prediction results and variables affecting the failure that match the thermodynamic modelling.

The contributions of this research are three folds:

1. Link is made between the generated XAI explanation to the thermodynamic performance modelling to discover the defective component which is still unexplored in previous researches.
2. The work applied a combination of explanation techniques from ML's aleatoric and epistemic uncertainty behaviour to global and local Shapley values contributions to ensure comprehensive explanation on prediction quality, model's optimization level, overall and individual prognostics.
3. The framework developed is a combination of all the mentioned error-proof methods, complementing each other weaknesses and taking advantage of their explanation abilities.

II. Literature Review

XAI is a field whose goal is to make the output mechanism of black-box ML and DL models transparent and understandable to humans [13]. XAI generates explanations of why an output is given by a model [14]. In the context of GT RUL estimation, XAI could potentially be used to explain which variables are responsible for failure, helping to confirm if the ML prediction matches the thermodynamic equations [15]. Subsequently, it could prove that the ML model's prediction is correct and logical.

[16] proposes QiTransformer architecture that integrates quantum-enhanced representation learning with causal inference and XAI to outperform traditional models on battery degradation and C-MAPSS aircraft engine datasets. [17] develops prognostic health indicators using wavelet denoising and auto-encoders combined with an ensemble of heterogeneous machine learning predictors (LSTM and GRU) and validates the approach using C-MAPSS data. Seeking to bridge the

 doi:10.20944/preprints202604.2033.v1

modality gap between sensor signals and text-based reasoning for machine health, [18] presents FD-LLM framework that aligns vibration signals with Large Language Models through string-based tokenization of spectra. To improve trustworthiness in safety-critical industrial prognostics by addressing the opacity of black-box models, [19] applies interpretable Concept Bottleneck Models (CBMs) that use component degradation modes as intermediate concepts for RUL prediction. Seeking to provide an adaptable and ethical predictive maintenance solution for complex industrial systems, [20] combines Autoencoder-based feature reduction with a CNN-LSTM network optimized by Particle Swarm Optimization (PSO) to improve accuracy and transparency in RUL prediction.

III. Methodology

In the development phase, data from an actual GT was initially collected. The variables corresponding to thermodynamic modelling were chosen and anomalous data were filtered from the dataset. Then, off-design healthy data modelling was performed, and data validation was executed. Some of the healthy data result, predicted from thermodynamic modelling, is then injected with degradation to create the failure data for training and testing. Next, Bayesian LSTM models were developed. A mix of healthy and failure predicted data obtained previously was prepared and divided for the LSTM model training and testing. Once training is done, the model performance was assessed, and the explanation model was developed from the Bayesian model where prediction explanation will be generated. Then, testing of healthy data was done using the trained model to establish the baseline HI, uncertainty level and explanation (Objective 1). In the evaluation phase, RUL prediction was executed with failure data. The prediction, and its uncertainty as well as the explanation were then compared with the baseline values to confirm anomaly occurrence (Objective 2). In the validation phase, the explanation was consulted and compared with thermodynamic modelling to confirm failure (Objective 3).

A. Data Description

Data from an 18.8 MW, twin-shaft industrial GT from an oil platform recorded over a one-year period, or 8737 hours, were used for preventive maintenance plan determination. The available parameters used in thermodynamic modelling and ML modelling are shown in Table I. Sensor reading errors were separated from the dataset and furthermore, the following values were adopted:

$\boldsymbol{P}_{\mathbf{amb}}$ = 101.3 kPa $\Delta\boldsymbol{P}_{\mathbf{1}}$ = 0.01 $\Delta\boldsymbol{w}_{\boldsymbol{bl}}$ = 0.06

ΔP_{CC} = 0.02 Δw_{bl-GGT}= 0.03 ΔP_{ed} = 0.025

Table I. Available Data for Modelling.

Input	Output	Purpose
T_{amb}, P_{amb}, PW_{C_ori}, N_{PT}	P_2, T_2, N_{GG}, T_4 and PW_{C_pred}	Thermodynamic Modelling
T_{amb}, P_2, T_2, N_{GG}, T_4	PW_{C_pred}	ML Model Modelling

B. Healthy Data Modelling

To estimate healthy data, thermodynamic modelling of the GT was performed. The schematic of the twinspool GT with free power turbine with the temperature, pressure and speed parameters are shown in Figure 1.

 doi:10.20944/preprints202604.2033.v1

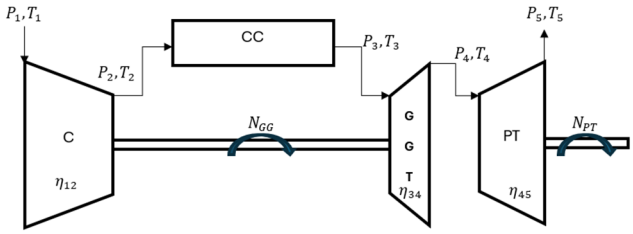


Figure 1. Twinspool gas turbine with free power turbine schematic.

The design points and off-design performance calculation following Tahan et al. [21] were adopted:

- The compressor power, PW_C, compressor inlet temperature, T_1, pressure, P_1, humidity and the power turbine speed, N_{PT} were the inputs to the model.
- The values of compressor inlet flow, w_1, compressor rotational speed, N_{GG}, gas generator turbine inlet temperature, T_3 and gas generator turbine pressure ratio, P_3/P_4 were estimated.

1. Calculate T_1 and P_1 with

$$T_1 = T_{amb} \tag{1}$$

$$P_1 = P_{amb} - \Delta P_1 * P_{amb} \tag{2}$$

2. With w_1, R_1, T_1, P_1 and N_{GG}, calculate w_{1cor} and N_{1cor}

$$N_{cor} = (N_{GG}/\sqrt{(T_1/Tsi)}) \tag{3}$$

$$w_{cor} = (w_1\sqrt{(T_1/Tsi)}/(P_1/Psi) \tag{4}$$

3. Use w_{1cor}, N_{1cor} and compressor curves to determine P_2/P_1 and η_{12}.
4. Calculate P_2 with

$$P_2 = P_1 \times P_2/P_1 \tag{5}$$

5. Calculate w_2 with

$$w_2 = w_1 - \Delta w_{bl} \tag{6}$$

6. Obtain T_2 using

$$T_2 = T_1 + T_1 \times \eta_{12}[\left(\frac{P_2}{P_1}\right)^{\left(\frac{\gamma_{12}-1}{\gamma_{12}}\right)} - 1] \tag{7}$$

where for calculating γ_{12}, the temperature is considered to be the mean of T_1 and T_2.

7. Calculate the compressor consumed power, PW_{C_pred} using

$$PW_{C\_pred} = w_1 \times cp_{12}(T_2 - T_1) \tag{8}$$

Where cp_{12} is calculated using the mean of T_1 and T_2 at constant pressure.

8. Use T_2, T_3, $T_3 - T_2$ and combustion charts to calculate the fuel flow, w_f.
9. Calculate P_3 with

$$P_3 = P_2 - \Delta P_{CC} * P_2 \tag{9}$$

10. Calculate w_3 with

$$w_3 = w_2 + w_f \tag{10}$$

11. Determine w_{3cor}, N_{3cor} using $w_3, R_3, T_3, P_3 and\ N_{GG}$.

$$N_{3cor} = (N_{GG}/\sqrt{T_3}) \tag{11}$$

$$w_{3cor} = (w_3\sqrt{T}_3/P_3) \tag{12}$$

 doi:10.20944/preprints202604.2033.v1

12. Use estimated P_3/P_4 and N_{3cor} to determine w_{3corc} and η_{34} employing the gas generator turbine.
13. Calculate T_4 with

$$T_4 = T_3 - T_3 \times \eta_{34}\left[1 - \left(\frac{P_4}{P_3}\right)^{\left(\frac{\gamma_{34}-1}{\gamma_{34}}\right)}\right] \quad (13)$$

where for calculating γ_{34} the temperature is considered to be the mean of T_3 and T_4.
14. Calculate PW_{GGT} with

$$PW_{GGT} = w_3 \times cp_{34}(T_3 - T_4) \quad (14)$$

where cp_{34} is calculated using the mean of T_3 and T_4 at constant pressure.
15. Calculate P_4 with

$$P_4 = P_3 \times P_4/P_3 \quad (15)$$

16. Calculate w_4 with

$$w_4 = w_3 + \Delta w_{bl-GGT} \quad (16)$$

17. Calculate w_{4cor}, N_{4cor} using w_4, R_4, T_4, P_4 and N_{PT}.

$$N_{4cor} = (N_{PT}/\sqrt{T_4}) \quad (17)$$

$$w_{4cor} = (w_4\sqrt{T}_4/P_4) \quad (18)$$

18. Calculate P_5 with

$$P_1 = P_{amb} + \Delta P_{ed} * P_{amb} \quad (19)$$

19. Using P_5/P_4 and N_{4cor} to determine w_{4corc} and η_{45} using the power turbine curve.
20. Calculate T_5 with

$$T_5 = T_4 - T_4 \times \eta_{45}\left[1 - \left(\frac{P_5}{P_4}\right)^{\left(\frac{\gamma_{45}-1}{\gamma_{45}}\right)}\right] \quad (20)$$

where for calculating γ_{45} the temperature is considered to be the mean of T_4 and T_5.
21. Calculate power generated by power turbine, PW_{PT} with

$$PW_{PT} = w_4 \times cp_{45}(T_4 - T_5) \quad (21)$$

where cp_{45} is calculated using the mean of T_4 and T_5 at constant pressure.

Component maps from Lazzaretto and Toffolo [22], scaled using from GasTurb software were chosen. The compressor, gas generator turbine and power turbine maps depicting corrected mass flow versus pressure ratio are applied as shown in Figure 2. The ß-line grid method was established to transmit the maps values into table forms [23].

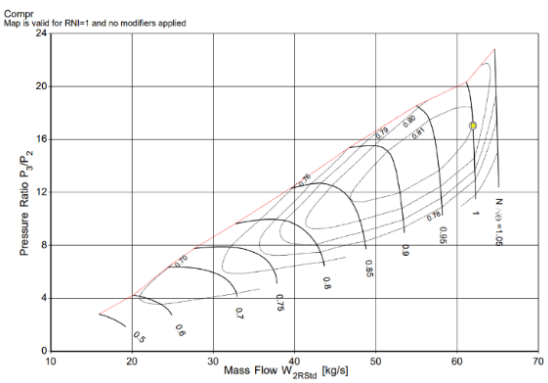


Figure 2. Compressor component map. (Lazzaretto & Toffolo 2001).

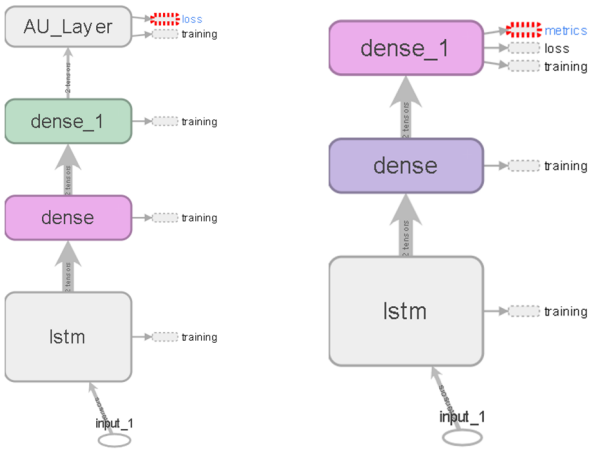
 doi:10.20944/preprints202604.2033.v1

C. Failure Data Creation

Once the thermodynamic modelling was done, fault was injected to the equations following the method from [24] to obtain the degradation data.

D. ML Models Development

Initially, a normal single LSTM layer with an FC layer f_x was developed. f_x was fed with normalized inputs and outputs of the system. The hyperparameters of f_x were obtained using Bayesian Hyperparameter Optimization. Next, the FC layer was replaced by probabilistic layers to build a Bayesian model, f_x^1 and its ensemble, f_x^2. The Bayesian version enables the quantification of all types of ML model's uncertainties. All the output layers are independent from one another, taking the LSTM output as input. The graphs are shown in Figure 3.



(a) (b)

Figure 3. Bayesian models structures (a) f_x^1 structure (b) f_x^2 structure.

- f_x^1 structure: As shown in Figure 3(a), this structure add aleatoric uncertainty (AU) layer to LSTM output. AU is uncertainty related to input data and it cannot be reduced with more data. The 'Dense' layer performs nonlinear operation. Then, 'Dense_1' layer, or fully connected layer, takes the input from the LSTM layer and predicts the mean and standard deviation of the prediction. The mean and standard deviation outputs are used later to construct the prediction distribution. The AU layer then samples from the predicted distribution.
- f_x^2 structure: As shown in Figure 3(b), this structure add epistemic uncertainty (EU) layer to LSTM output. EU is uncertainty related to model's parameters and it can be reduced with more data. The weights from f_x^1 are transferred to a similar model (f_x^2) without AU layer. 9 copies of f_x^2 with similar parameters are created, forming an ensemble. The copies are initialized randomly with random seeds, trained and tested. EU can be obtained by monitoring the predictions discrepancies of the 10 ensemble models (f_x^2 and its copies).

E. Models' Training and Prediction

For this modelling, N_{PT} is excluded as it does not affect PW_{C_pred} measurement. Firstly, the erroneous measures were removed from the data. Then, some data was discarded as they are outside of the 55% to 83% of the gas turbine rated power and will not be helpful for the generalization of the ML model after training. Next, the cleaned data was split into training, validation, and testing datasets. Each sequence of inputs and output of the model were set to 1 week or 168 hours (14 days ahead forecast). The input data was normalized using min-max normalization. HI was created from

 doi:10.20944/preprints202604.2033.v1

PW_{C_pred} measurement. For f_x^1, the training loss metric is the Negative Log Likelihood (Negloglik) loss while for f_x^2, is the root mean square error (RMSE) loss.

F. Multi-Step Ahead, Long Term Forecast

To assess the performance of the model across different literature treating long term forecasting, MAPE metric is used. MAPE is independent from the scale of the measurement, allowing the comparison across different benchmark data. [25] reported the criteria below in Table II. to evaluate MAPE metric in a general prediction problem.

Table II. MAPE Value and Prediction Performance Classification.

Input Data Hour	Input Data State
<10	Highly accurate prediction
10–20	Good prediction
20–50	Reasonable prediction
>50	Inaccurate prediction

The task undertaken in this work is a 14-days in advance prediction problem and the ML prediction is given in hours. Thus, in reality, it is a 336-hours (or steps) ahead forecasting task. According to [26] the problem can be classified as a multi-step ahead, long-term forecasting problem. Thus, we do not expect the performance of the model to be similar to short-term forecasting.

G. Explanation Mechanisms

Uncertainty trend, range and SHAP explanation were employed as the explanation mechanisms:

- Uncertainty trend is exploited to explain the model's confidence when it predicts certain output.
- Apart from the uncertainty behavior, a quantitative approach to support the prediction is vital. For failure data, the uncertainty bounds at 95% confidence interval (CI) need to be strictly below the healthy data region for clear decision making. The upper and lower bounds can be estimated at 95% confidence interval (CI) using equations (22) and (23).

$$\text{Upper bound} = \mu + Z * (\sigma/\sqrt{\bar{N}}) \tag{22}$$

$$\text{Lower bound} = \mu - Z * (\sigma/\sqrt{\bar{N}}) \tag{23}$$

- Meanwhile SHAP explanation is used to highlight the features responsible for RUL estimation.

H. Shapley Value Explanation

SHAP is a game theoretical approach to explain the output of any machine learning model. It evaluates the contribution of each feature to the prediction by using Shapley values. SHAP can be both global and local explainability approaches. Shapley values determine the importance of a single feature by considering the outcome of each possible combination of features. In other words, the Shapley value is the average expected marginal contribution of a feature across all possible combination of features.

The formulation of the Shapley value of feature j are given by

$$\phi_j(val) = \sum_{\mathcal{E}\subseteq\{x_1,\ldots,x_p\}\setminus\{x_j\}} \frac{|\mathcal{E}|!\,(p - |\mathcal{E}| - 1)!}{p!} (val(\mathcal{E} \cup \{x_j\}) - val(\mathcal{E}) \tag{24}$$

$$val_x(\mathcal{E}) = \int \hat{f}(x_1,\ldots,x_p) d\mathbb{P}_{x\notin\mathcal{E}} - E_X(\hat{f}(X) \tag{25}$$

 doi:10.20944/preprints202604.2033.v1

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z_j' \tag{26}$$

The SHAP framework applies cooperative game theory to decompose a machine learning model's output into the input features' individual contributions. (25) defines the Shapley value, ϕ_j as the fair credit assigned to a feature. This is calculated by averaging its marginal contribution, the difference in the value function between a set containing the feature of interest; the feature whose contribution is calculated, $val(\mathcal{E} \cup \{x_j\})$ and one without it, $val(\mathcal{E})$, weighted, $\frac{|\mathcal{E}|!(p-|\mathcal{E}|-1)!}{p!}$ across all possible feature combinations, $\mathcal{E}$. $\mathcal{E}$ is a subset of the total p features and x is the instance's vector to be explained. $val_x(\mathcal{E})$represents the expected prediction given a specific subset of features compared against the overall expected value, $E_X(\hat{f}(X))$ of the model. (26) incorporates these components into a linear explanation model, $g(z')$, where the final prediction is expressed as the base value, ϕ_0 plus the sum of the quantified impacts of all active features, $\sum_{j=1}^{M} \phi_j z_j'$. ϕ_0 is the average predictions of the entire dataset.$\sum_{j=1}^{M} \phi_j z_j'$ represents the total individual feature contribution in a prediction. Shapley value, ϕ_j thus moves the prediction positively or negatively from that average prediction. $z' \in \{0,1\}^M$ describes the presence of interested features in the feature's combination with $z' = 0$ means the interested feature is absent in the combination and $z' = 1$ signifying the feature is present. M is the maximum coalition size and $\phi_j \in R$ is the Shapley values for a feature j.

SHAP can generate global, or overall explanation through and local, or individual sequence explanations. However, it is not compatible with probabilistic DL and only accepts a single output vector for explanation. Thus, a workaround, in the form of a non-probabilistic model labelled as f_x^3, was developed as shown in Figure 4. Note that f_x^3 has the same layers and weights as those figured along the explanation path in f_x^1, except the weights in dense2 of f_x^1. Here, only the weights corresponding to the mean were used and transferred to f_x^3 while the weights associated with the standard deviation were ignored. The output layer out3 in f_x^3 sliced only the first value of each sequence vector and arranged them in a single vector for the explanation.

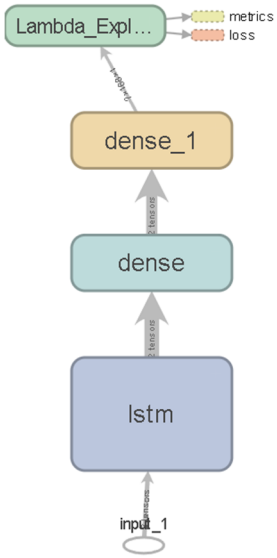


Figure 4. Explanation models architecture, f_x^3.

In order to define the HI baseline, outlier detection using interquartile (IQR) range is employed [27]. The first quartile ($Q1$), median and third quartile ($Q3$) are calculated from the predicted thermodynamic data. Then the formula to define the upper and lower anomaly thresholds are performed as follows:

 doi:10.20944/preprints202604.2033.v1

$$IQR = Q1 - Q3 \quad (27)$$

$$\text{Lower anomaly threshold} = Q1 - 1.5 \times IQR \quad (28)$$

$$\text{Upper anomaly threshold} = Q3 + 1.5 \times IQR \quad (29)$$

The model was trained using a mix of majority healthy and minority failure data to mimic the actual condition as closely as possible. To monitor the evolution of the forecast performance and the explanation as the prediction approaches the final failure date, 2 cases were considered.

Firstly, the model was tested with healthy testing data to establish the healthy AU level and EU level while the global explanation is taken as baseline explanation. The first failure scenario was using partial failure data from the 0th to 168th hours in week 1 corresponding to the combination of healthy and failure data to predict the 336th to 504th hours in week 3 corresponding to healthy and failure output data. The second failure scenario was using the 96th to 264th hours in week 1 and 2 corresponding to fully failure data to predict the 432nd to 600th hours in week 3 and 4 corresponding to failure output data as shown in Table III.

Table III. Prediction with Failure Data.

Input Data Hour	Input Data State	14 Days Ahead Output Data Hour	Output Data State
0 to 168 (Week 1)	Healthy and Failure	336-504 (Week 3)	Healthy and Failure
96 to 264 (Week 1 and 2)	Failure	432 to 600 (Week 3 and Week 4)	Failure

Anomaly can be detected by comparing the AU level with AU baseline. If failure prediction is lower than baseline AU covers mostly the failure level, anomaly possibility high. The difference between global explanation and the baseline explanation can also indicate anomaly. The global explanation of the failure then be compared with the thermodynamic modelling, confirming if the model predicts correctly. The local explanation, on the other hand, can explain the evolution of variables influencing the failure prediction.

IV. Results and Discussion

A. Thermodynamic Healthy Data Modelling

The MAPE results for the health parameters are shown in Table IV. The modelling shows high accordance between the actual and predicted data like the example shown in Figure 5, with average MAPE of 1.13. The average result shows high compatibility between the actual and predicted data.

 doi:10.20944/preprints202604.2033.v1

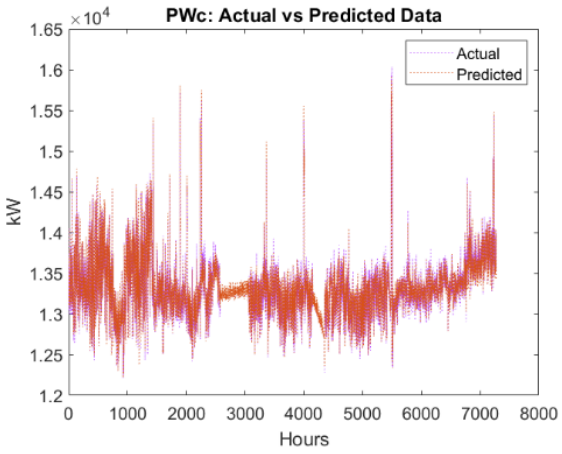


Figure 5. Predicted PW_{C_pred} vs PW_C using thermodynamic modelling.

Table IV. MAPE Results of Thermodynamic Modelling.

Parameter	P_2	N_{GG}	T_4	PW_{C_pred}	PW_{GGT}	*Average*
MAPE (%)	0.71	1.32	2.13	0.40	1.12	1.14

Additionally, the average result comparison compared to other published literature on twinspool GT [28,29] is presented in Table V. The results obtained showed that the modelling is at par with other published results.

Table V. MAPE Results Comparison with Other Works.

Work	**Current Work**	Hu et al. 2022	Sankar et al. 2022
MAPE (%)	**1.13**	1.50	2.72

B. Anomaly Threshold Values

Threshold values and the equivalent HI calculated from the predicted thermodynamic data using (28) and (29) are presented in Table VI.

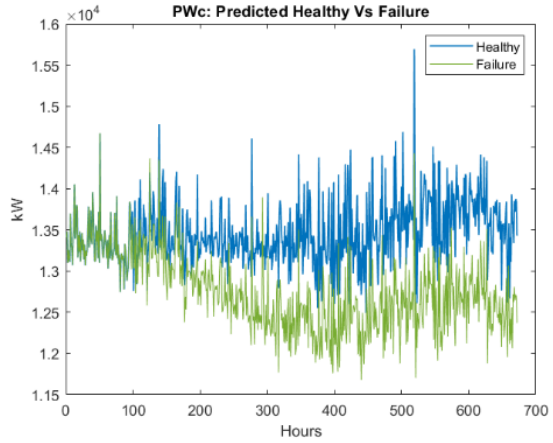
Table VI. Anomaly Thresholds.

Threshold	**Equivalent HI**
Lower = 12707.78 kW	0.37
Upper = 13882.48 kW	0.78

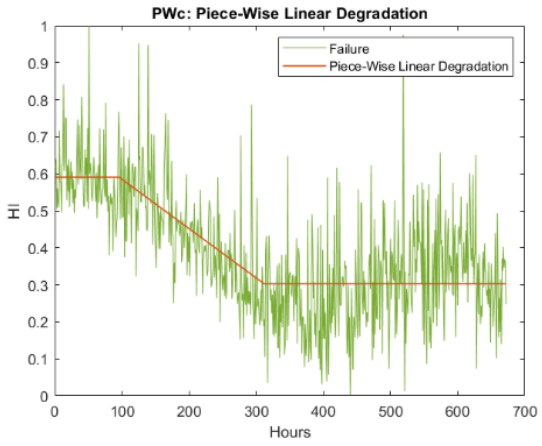
C. Multi-Step Ahead, Long-Term Forecasting

In this forecasting mode, the prediction needs to be compared with the trend line that best fits the actual data. The important indicators for decision making are the HI trend as well as the average level of the future life. Failure was injected incrementally from 96th hour to 312th hour like shown in Figure 6(a). This indicates a piece-wise linear degradation [30,31]. Figure 6(b) shows an example of the fitting results for the true failure HI with piece-wise linear degradation by taking the average HI value.

 doi:10.20944/preprints202604.2033.v1



(a)



(b)

Figure 6. Piece-wise linear degradation Health Index.

As can be seen in Table VII, the predictions reside in the good performance area according to Table II. To give an idea of the performance of the model compared to other published works, comparison in Table VIII is done.

Table VII. MAPE Results of The DL Predictions.

Scenario	Healthy Data with AU	Healthy Data with EU	Full Failure Data with AU
MAPE (%)	18.04	4.37	18.13

Table VIII. Anomaly Thresholds.

Work	Dataset	Maximum Prediction Horizon	MAPE (%)
Feng et al. 2020	Wind Speed	5 steps	5-15
Nguyen et al. 2021	Reactor Coolant Pumps	18 steps	13-30
Current Work	Gas Turbine	336 steps	18

 doi:10.20944/preprints202604.2033.v1

The prediction horizon used in this work is far superior compared to the two others. Observing these results and considering the long term prediction horizon, we can consider that the performance of the model resides between good to highly accurate.

D. Healthy Data Predictions with Uncertainty Explanation

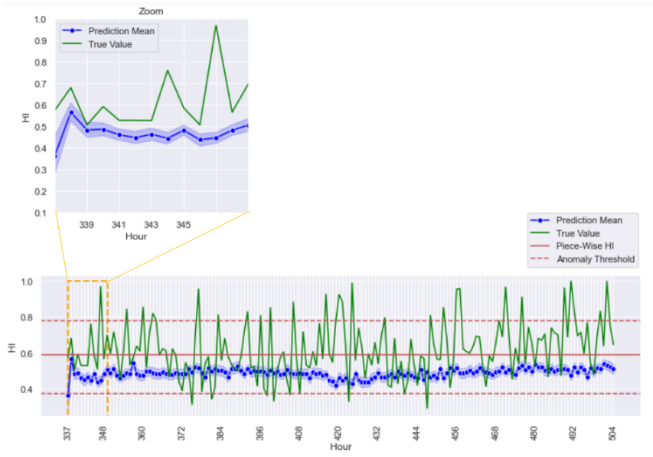
Table IX lists the HI and uncertainty levels of healthy data. The initial predicted HI serves as baseline value for healthy data prediction.

Table IX. Health Index and Uncertainty Level For Healthy Data.

True HI	HI with AU	AU Level	HI with EU	EU Level
0.591	0.487	0.022	0.565	0.002

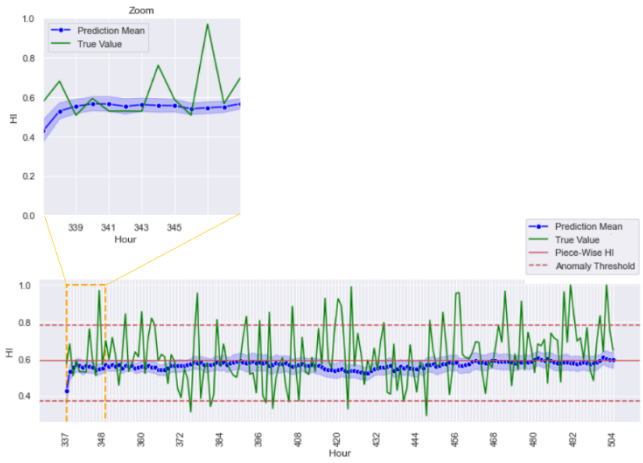
Healthy Data Uncertainty Explanation:

- As can be seen from Figure 7(a) and 7(b), the predictions with AU and EU boundaries are within the healthy data range. The possibility of healthy data prediction is thus high.
- The AU uncertainty bounds considering 95% CI, shows that the sampled values are located +/- 0.07% from the mean HI. This low AU level infers that the testing data is nearly similar to the training data and the model is confident in its prediction.
- The EU uncertainty bounds considering 95% CI, shows that the sampled values are located +/- 0.06% from the mean HI. The very low EU level indicates that the developed model represents well the data.
- Additionally, the low uncertainty levels imply that the prediction and explanation generated from this model can be trusted.



(a)

 doi:10.20944/preprints202604.2033.v1



(b)

Figure 7. Healthy data prediction (a) Prediction with aleatoric uncertainty (b) Prediction with epistemic uncertainty.

E. Failure Data Predictions with Uncertainty Explanation

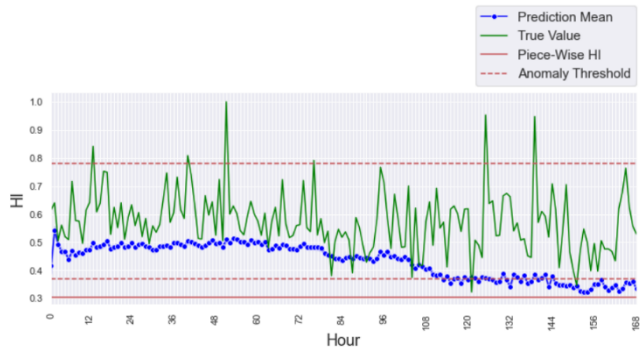
The failure prediction and uncertainty are analyzed against both the present and future ground truth HI. On one hand, comparing the prediction with the present values enables us to evaluate the remaining useful life between the present and predicted target data. This is also the only view a maintenance personnel can appreciate due to the absence of future values. On the other hand, comparing the prediction to the future values enables us to compare the worst predicted degradation level to the true future degradation level.

Table X lists the HI and uncertainty levels of failure data.

Table X. Health Index and Uncertainty Level For Failure Data.

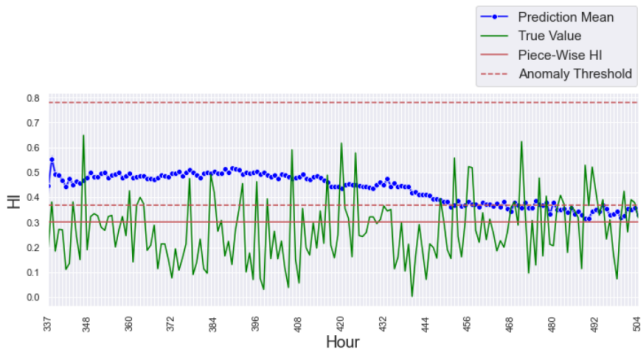
True HI	HI with AU for Partial Failure	AU Level for Partial Failure	HI with AU for Full Failure	AU Level for Full Failure
0.591	0.431	0.146	0.474	0.147

The partial and full failure prediction compared to true present and future values are shown in Figure 8(a) and 8(b) and Figure 9(a) and 9(b) respectively.



(a)

 doi:10.20944/preprints202604.2033.v1



(b)

Figure 8. Partial failure prediction (a) Compared to present data (b) Compared to future data.

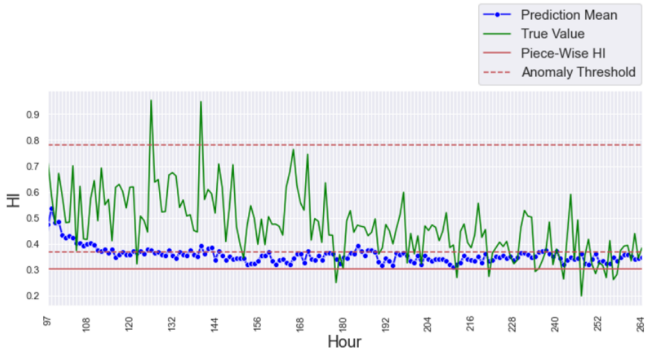
Partial Failure Uncertainty Explanation:

- For partial failure, the prediction with AU boundaries is within the healthy data range from beginning until around 106th and 444th hour respectively. This indicates that the possibility of healthy data prediction is high initially when the healthy part of the test data is used.
- Then, the prediction passed to the degradation range around 120th and 456th hour respectively until the end, pointing to the high possibility of failure when the failure part of the test data is used.
- The final degradation level's prediction in Figure 8(b) also approaches the true future values.
- The average AU is higher than the healthy data AU level of 0.022 in Table VIII which points to anomaly.
- The AU uncertainty bounds considering 95% CI, shows that the sampled values are located +/- 0.51% from the mean HI. The low level infers that the testing data is nearly similar to the training data and presents a good prediction confidence.

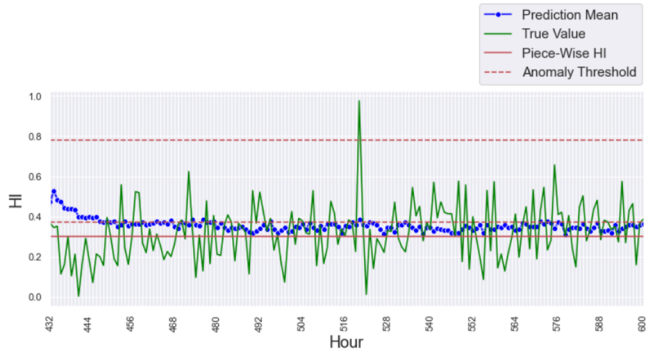
Full Failure Uncertainty Explanation:

- The predictions are within the healthy data range from beginning until around 106th and 444th hour respectively and then descended rapidly to the anomaly data range around 120th and 456th hour respectively until the end, pointing to the high possibility of failure.
- The final degradation level's prediction in Figure 9(b) also approaches the true future values.
- For full failure, the average AU which is 0.147 is higher than the healthy AU. This is an indication that the failure data nature is different from the healthy training data and that anomaly has occurred.
- The AU uncertainty bounds considering 95% CI, show that the sampled values are located +/- 0.62% from the mean HI. The low level infers that the testing data is nearly similar to the training data and presents a good prediction confidence.
- In Figure 9(a), the AU bounds at 95% CI is mostly always below the present true HI level. The possibility of degradation is thus high.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 April 2026 doi:10.20944/preprints202604.2033.v1



(a)



(b)

Figure 9. Full failure prediction (a) Compared to present data (b) Compared to future data.

F. Healthy and Failure Data SHAP Explanation

Healthy Data SHAP Global Explanation: The SHAP global explanation for healthy data is presented in Figure 10:

- P_2, T_1, N_{GG}, T_4, and T_2 contribute to PW_{C_pred} prediction according to order of contribution.
- The contributions are almost balanced in both directions, indicating equal total forces that pull the prediction up (healthy side) or pushing it down (failure side).

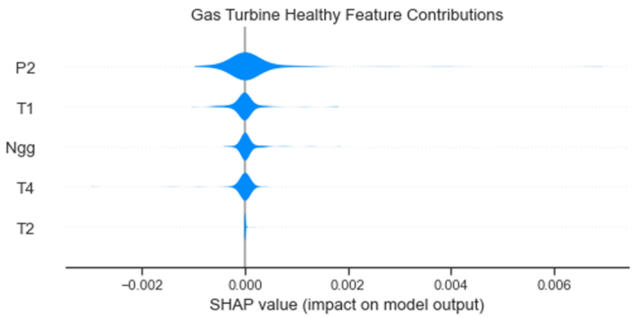


Figure 10. Global explanation for healthy data modelling.

This explanation is compatible with thermodynamic modelling measures and constitutes the baseline explanation for healthy data prediction.

The SHAP local explanation for healthy, partial failure and full failure data are presented in the heatmap Figure 12.

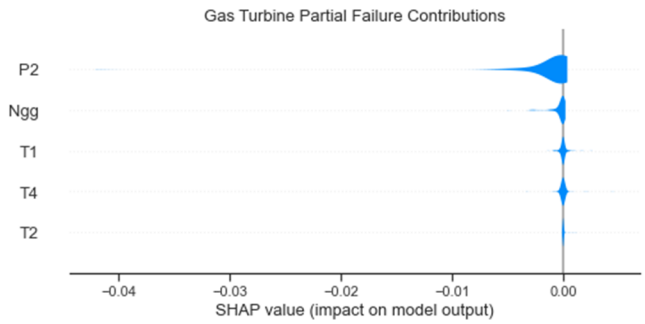
 doi:10.20944/preprints202604.2033.v1

Healthy Data Local Explanation: During this state, features are very volatile, frequently swapping both their ranking and polarity:

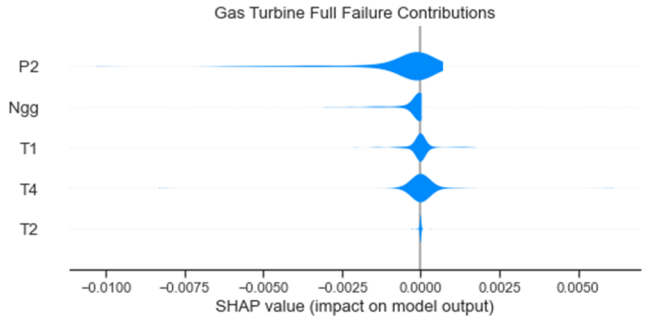
- From 1st to 10th hour, $-T_2$ and $+T_1$ trade the most significant contribution spot. Then $+P_2$ trades places with $-T_1$ and $+N_{gg}$ for second and third contributors.
- From 11th to 166th hour, the system enters a period of high volatility where $-P_2$,$+T_1$, and $-T_1$ frequently swap the top two rankings. Meanwhile, $-T_2$ remains the fifth contributor.
- The constant polarity shifts and contribution rank trades suggest a system in operational equilibrium which is typical of a healthy state.

Failure Data Global Explanation: The SHAP global explanation for the partial and full failure data are presented in Figure 11 respectively:

- For both failure modes, P_2 contributes the most to PW_{c_pred} prediction followed by N_{GG},T_1, T_4 and T_2. This explanation is thus different from the healthy data explanation which points to anomaly.
- For both failure modes the contribution of strongest variables are toward negative directions, which indicates degradation.
- For partial failure, P_2 and N_{GG} contributions level increase significantly as data evolves from healthy to failure. $N_{GG}$and T_4 contributions also increase slightly. This is because it takes more contributions to change the state from healthy to failure.
- For full failure, P_2 and N_{GG} contributions level decrease significantly as data is completely in failure status. $N_{GG}$and T_4 contributions also decrease slightly. This is because it does not take too much contributions to preserve the failure state.



(a)



(b)

Figure 11. Global explanation for failure prediction (a) Partial failure (b) Full failure.

 doi:10.20944/preprints202604.2033.v1

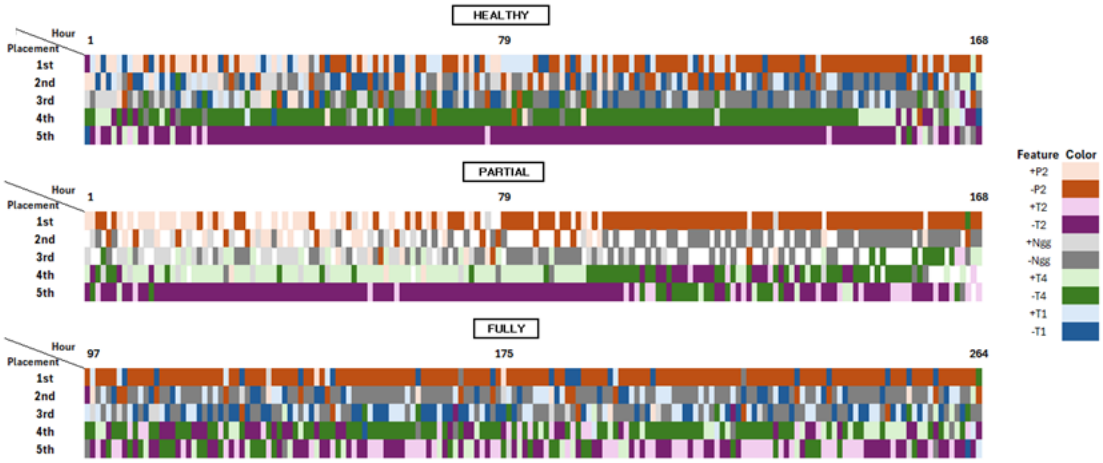


Figure 12. Local explanation heatmap.

This explanation is compatible with thermodynamic modelling.

Partial Failure Local Explanation: This state shows an initial equilibrium that converges into negative dominance.

- Before 96th hour, the 1st place position is where $-P_2$ and $+/-T_1$ frequently trade places.
- $-N_{gg}$ and T_1consistently occupy the 2nd and 3rd spots. Their polarities shift often, suggesting the system is attempting to balance fluctuations within the partial failure state.
- Throughout the entire 96-hour sequence, $-T_2$ remains almost exclusively in the 5th place ranking.
- After 96th hour, the contribution trend becomes more fixed and is defined by the consistent dominance of negative features.
- The 1st place position settles almost permanently into $-P_2$. Unlike before, there are far fewer instances where $+P_2$ and $+T_1$ challenge for the top spot.
- $-N_{gg}$ transitions from a rotating secondary feature to the nearly permanent 2nd place contributor. This duo of $-P_2$ and -N_{gg} defines the late-cycle trend.
- T_1, which was highly influential before 96th hour, begins to drop in the rankings. It moves from the top three to the 4th position and exhibits polarity flips, such as the shift to $+T_1$ in the final hours.
- The trend before 96th hour represents a fluctuating partial failure where positive and negative influences from pressure and ambient temperature compete. After 96th hour, the trend transitions into a consistent downward decline led by negative pressure and gas generator speed, with other features losing their relative importance.

Full Failure Local Explanation: This state exhibits the most stable negative contributions in its top ranking until the end sequence:

- From 1st to 167th hour, $-P_2$ contributes the most with a dominant negative contribution. Then $-N_{gg}$ trades places with $+T_1$ and $-T_1$ for second and third contributors, though $-N_{gg}$ maintains a much more stable presence in the top three.
- Throughout the sequence, $-T_4$ and $-T_2$ trade third and fourth places with consistent negative contributions.
- The stable negative contributions across pressure, speed, and temperature consistently pull the prediction values lower as the system fails.

G. Explanation Validation with Thermodynamic Equations

In the healthy data modelling, the explanation in Figure 10 shows that:

 doi:10.20944/preprints202604.2033.v1

- P_2, T_1, N_{GG}, T_4 and T_2 contribute to PW_{C_pred} prediction according to order of contribution. This explanation is compatible with the thermodynamic modelling. The contributions are almost all balanced in both directions.
- In terms of thermodynamic modelling equation, P_2 is proportionate to T_2 (7) and cp_{12} values which greatly influenced PW_{C_pred} (8).
- T_1 is proportionate to P_2/P_1 (3) and compressor map) and T_2 which influenced PW_{C_pred}.
- N_{GG}, influences P_2/P_1 and η_{12} (3) and compressor map) which in turn influence T_2 and PW_{C_pred}.
- T_4, on the other hand, is inversely proportionate to cp_{34} and PW_{GGT} (14) which is conditioned by PW_{C_pred}.
- However, T_4 influence on PW_{C_pred} is lesser compared to T_1, P_2 and N_{GG} as there is no direct relationship between PW_{GGT} and PW_{C_pred}. T_2 is just a secondary product of T_1, P_2 and N_{GG} (7), thus its contribution to PW_{C_pred} is weaker than those features.

In failure prediction, the global explanation in Figure 11 show that:

- P_2, N_{GG}, T_1, T_4 and T_2 contribute to PW_{C_pred} prediction according to order of contribution. Again, the explanation is compatible with the thermodynamic modelling.
- The global explanation is different from the healthy data explanation, which points to anomaly.
- The contribution of strongest variables are toward negative directions, which also indicates anomaly.
- It also implies that the compressor discharge health parameter, P_2 and N_{GG} as the most contributing anomaly features, while T_4 and T_2 contributions had fallen to 4th and 5th place. This implies an anomaly coming from the compressor.
- Since PW_{C_pred} values decrease over time, the only plausible cause is the decrease of T_2 which increases γ_{12} and reduces cp_{12} over time (8).
- T_2 reduction, in turn, can only be caused in this case by reduction in compressor pressure ratio, P_2/P_1 (7), as compressor efficiency is not part of the ML modelling.
- P_2/P_1 reduction is of course proportionate to P_2 and N_{GG} reduction (3) and compressor map).
- Since PW_{GGT} is conditioned by PW_{C_pred}, T_4 will also have to change (14).
- T_1, while having influence on T_2, does not change due to failure injection, thus the weaker contribution compared to P_2 and N_{GG}.

The health parameters' pattern points to compressor failure, where P_2 and N_{GG} decrease, which reduces T_2, which in turn lowered PW_{C_pred} value (8) and T_4 measures (14). According to local explanations in Figure 12, features contributions are negative, pulling the prediction lower. This indication strengthens the information obtained from the AU level and behavior as well as the predicted HI level.

The transparency of XAI based ML model, which is equivalent to those present in thermodynamic modelling, increases the confidence in ML model utilization. Additionally, by knowing in advance, the components responsible for future failure, maintenance activity and resources can be better organized and optimized. In this case study, it is obvious that the health of the compressor needs to be checked first as the first two most contributing features to failure prediction are P_2 and N_{GG}.

V. Conclusion

The goal of this work is to develop an uncertainty aware explainable AI machine learning (ML) model that can predict and confirm the occurrence of gas turbine (GT) failure for preventive maintenance planning. Specifically, Bayesian ML models with (SHAP) explanation capable of life prediction with uncertainty and generating explanation are developed. A 1-year worth of twinspool GT data recording was used for this end. The healthy GT data was estimated using thermodynamic modelling that shows high compatibility with the recorded data. The healthy data was injected with fouling failure and the resultant data was used to train and test the ML model.

 doi:10.20944/preprints202604.2033.v1

A 14-days in advance prediction was executed using healthy and failure testing data. The low uncertainties show that the developed model represents well the data. Consequently, the prediction and explanation of this model can be trusted. The failure life predictions point to anomaly. The uncertainty level shows high possibility of either degraded or failure state, strengthening the life prediction belief.

Consequently, the low and decreasing uncertainty indicator leads to accurate explanations which are compatible with the thermodynamic modelling and point to compressor failure. The explanation points mainly to compressor discharge pressure and gas generator speed reduction which cause decrease in compressor outlet temperature, output power and combustor outlet temperature and all these features contribute negatively to the prediction.

Nomenclature

	Description	***Unit***
N	Rotational speed	(RPM)
P	Pressure	(kPa)
PW	Total power output	(MW)
T	Temperature	(°K)
$\boldsymbol{cp}$	Specific heat	(kJ/ kg-K)
$\boldsymbol{w}$	Mass flow rate	(kg/hr)
$\boldsymbol{\Delta P}$	Pressure loss	(kPa)
$\boldsymbol{\Delta w}$	Air bleed	(kg/h)
Greek Letter		
$\boldsymbol{\eta}$	Efficiency	%
$\boldsymbol{\gamma}$	Isentropic index	
Superscript		
Bl	Blowoff	
C	Compressor	
CC	Combustion chamber	
Cor	Corrected	
Corc	Corrected using curve	
Ed	Exhaust duct	
GG	Gas generator	
GGT	Gas generator turbine	
PT	Power turbine	
amb	Ambient condition	
1	Compressor inlet	
2	Compressor outlet	
3	Combustion chamber outlet	
4	Gas generator outlet	
5	Power turbine outlet	

 doi:10.20944/preprints202604.2033.v1

References

1. C. Yu et al., "Study on the effects of nanobubble-enhanced diesel spray technology on the performance of heavy-duty gas turbines," Thermal Science and Engineering Progress, vol. 73, p. 104677, May 2026. doi:10.1016/j.tsep.2026.104677.
2. R. Hwang, J. Lee, J. Kim, I. Moon, and M. Oh, "Autonomous Digital Twin Framework for gas turbine combined cycle control loops: Comparative study of proportional-integral control, reinforcement learning, and reinforcement learning with agents," Energy and AI, vol. 24, p. 100727, May 2026. doi:10.1016/j.egyai.2026.100727
3. B. S. Mohd Irwan Shah, A. J. Ishak, M. K. Hassan, and N. M. Norsahperi, "Revolutionizing gas turbine performance analysis with Deep Learning Powered Digital Twin," e-Prime – Nexus of Electrical, Electronic, and Intelligent Engineering, vol. 17, p. 201178, Sep. 2026. doi:10.1016/j.eprime.2026.201178
4. J. Zeng and Z. Liang, "Predictive group maintenance using probabilistic prognostics and deep reinforcement learning," Computers & Industrial Engineering, vol. 212, p. 111738, Feb. 2026. doi:10.1016/j.cie.2025.111738
5. Ayman, A. Onsy, O. Attallah, H. Brooks, and I. Morsi, "Feature learning for bearing prognostics: A comprehensive review of machine/Deep learning methods, challenges, and opportunities," Measurement, vol. 245, p. 116589, Mar. 2025. doi:10.1016/j.measurement.2024.116589
6. R. Machlev, "Ev battery fault diagnostics and Prognostics using Deep Learning: Review, Challenges & Opportunities," Journal of Energy Storage, vol. 83, p. 110614, Apr. 2024. doi:10.1016/j.est.2024.110614.
7. W. Cheng et al., "Diagnostics and Prognostics in power plants: A systematic review," Reliability Engineering & System Safety, vol. 255, p. 110663, Mar. 2025. doi:10.1016/j.ress.2024.110663
8. C. Regazzoni, A. Krayani, G. Slavic, and L. Marcenaro, "Probabilistic anomaly detection methods using learned models from time-series data for multimedia self-aware systems," Advanced Methods and Deep Learning in Computer Vision, pp. 449–479, 2022. doi:10.1016/b978-0-12-822109-9.00022-9
9. P. Schummer et al., "Machine learning-based network anomaly detection: Design, implementation, and evaluation," AI, vol. 5, no. 4, pp. 2967–2983, Dec. 2024. doi:10.3390/ai5040143.
10. M. Z. Naser, "Fundamental flaws of physics-informed neural networks and explainability methods in Engineering Systems," Computers & Industrial Engineering, vol. 212, p. 111704, Feb. 2026. doi:10.1016/j.cie.2025.111704.
11. X. Yuan, T. Bai, and C. Peng, "Hybrid modeling method for reactor coolant loop combining data-driven and physics-based constraints," Energy, vol. 345, p. 140177, Feb. 2026. doi:10.1016/j.energy.2026.140177.
12. Ensemble Modeling with a Bayesian Maximal Information Coefficient-Based Model of Bayesian Predictions on Uncertainty Data
13. Explainable AI for industrial fault diagnosis: A systematic review.
14. W. Yang et al., "Survey on Explainable AI: From Approaches, Limitations and Applications Aspects", *Hum-Cent Intell Syst*, vol. 3, no. 3, pp. 161–188, Aug. 2023, doi: 10.1007/s44230-023-00038-y.
15. S. A. and S. R., "A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends", *Decision Analytics Journal*, vol. 7, p. 100230, Jun. 2023, doi: 10.1016/j.dajour.2023.100230
16. M. Gandhudi, A. P.J.A., S. Srinivas, and G. G.R., "Causal inference and explainable artificial intelligence based quantum deep learning for remaining useful lifetime prediction," Knowledge-Based Systems, vol. 340, p. 115669, May 2026. doi:10.1016/j.knosys.2026.115669.
17. M. Soualhi, K. T. P. Nguyen, and K. Medjaher, "Explainable RUL estimation of turbofan engines based on prognostic indicators and heterogeneous ensemble machine learning predictors," Engineering Applications of Artificial Intelligence, vol. 133, p. 108186, Jul. 2024. doi:10.1016/j.engappai.2024.108186.
18. H. Qaid et al., Large language models for explainable fault diagnosis of machines, 2025. doi:10.2139/ssrn.5404598.
19. F. Forest, K. Rombach, and O. Fink, "Interpretable prognostics with concept bottleneck models," Information Fusion, vol. 124, p. 103427, Dec. 2025. doi:10.1016/j.inffus.2025.103427.

 doi:10.20944/preprints202604.2033.v1

20. Priyadarshini, "An explainable Autoencoder-based feature extraction combined with CNN-LSTM-PSO model for improved predictive maintenance," Computers, Materials & Continua, vol. 83, no. 1, pp. 635–659, 2025. doi:10.32604/cmc.2025.061062.
21. M. Y. Razak, Industrial Gas Turbines: Performance and Operability. Boca Raton, Cambridge, England: CRC Press ; Woodhead Pub, 2008.
22. Lazzaretto and A. Toffolo, "Analytical and Neural Network Models for Gas Turbine Design and Off-Design Simulation," *International Journal of Thermodynamics*, vol. 4, no. 4, pp. 173–182, Dec. 2001, doi: https://doi.org/10.5541/ijot.1034000078.
23. M. Razmjooei, F. Ommi, and Z. Saboohi, Experimental Analysis and modeling of gas turbine engine performance: Design Point and off-design insights through system of equations solutions, 2024. doi:10.2139/ssrn.4823484.
24. Zwebek and P. Pilidis, "Degradation effects on combined cycle power plant performance: Part 1 — gas turbine cycle component degradation effects," Volume 2: Coal, Biomass and Alternative Fuels; Combustion and Fuels; Oil and Gas Applications; Cycle Innovations, Jun. 2001. doi:10.1115/2001-gt-0388.
25. E. Vivas, H. Allende-Cid, and R. Salas, "A systematic review of statistical and Machine Learning Methods for electrical power forecasting with reported MAPE score," Entropy, vol. 22, no. 12, p. 1412, Dec. 2020. doi:10.3390/e22121412.
26. S. Suradhaniwar, S. Kar, S. S. Durbha, and A. Jagarlapudi, "Time series forecasting of Univariate Agrometeorological Data: A comparative performance evaluation via one-step and multi-step ahead forecasting strategies," Sensors, vol. 21, no. 7, p. 2430, Apr. 2021. doi:10.3390/s21072430.
27. Ch. S. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, "An outliers detection and elimination framework in classification task of Data Mining," Decision Analytics Journal, vol. 6, p. 100164, Mar. 2023. doi:10.1016/j.dajour.2023.100164.
28. B. Sankar, B. J. Shah, S. Jana, R. K. Satpathy, and G. Gouda, "Modeling of degradation in gas turbine engine by modified off design simulation," Defence Science Journal, vol. 72, no. 2, pp. 135–145, May 2022. doi:10.14429/dsj.72.15428.
29. M. Hu et al., "Digital Twin Model of gas turbine and its application in warning of performance fault," Chinese Journal of Aeronautics, vol. 36, no. 3, pp. 449–470, Mar. 2023. doi:10.1016/j.cja.2022.07.021.
30. Asif et al., "A deep learning model for remaining useful life prediction of aircraft turbofan engine on C-MAPSS dataset," IEEE Access, vol. 10, pp. 95425–95440, 2022. doi:10.1109/access.2022.3203406.
31. Asif et al., "A deep learning model for remaining useful life prediction of aircraft turbofan engine on C-MAPSS dataset," IEEE Access, vol. 10, pp. 95425–95440, 2022. doi:10.1109/access.2022.3203406.