ROYAL AERONAUTICAL SOCIETY

RESEARCH ARTICLE

Gas turbine prognostics via Temporal Fusion Transformer

A.D. Fentaye and K.G. Kyprianidis

Department of Sustainable Energy Systems, Mälardalen University, Västerås, Sweden
Corresponding author: A. D. Fentaye; Email: amare.desalegn.fentaye@mdu.se

Abstract

Gas turbines play a vital role in various industries. Timely and accurately predicting their degradation is essential for efficient operation and optimal maintenance planning. Diagnostic and prognostic outcomes aid in determining the optimal compressor washing intervals. Diagnostics detects compressor fouling and estimates the trend up to the current time. If the forecast indicates fast progress in the fouling trend, scheduling offline washing during the next inspection event or earlier may be crucial to address the fouling deposit comprehensively. This approach ensures that compressor cleaning is performed based on its actual health status, leading to improved operation and maintenance costs. This paper presents a novel prognostic method for gas turbine degradation forecasting through a time-series analysis. The proposed approach uses the Temporal Fusion Transformer model capable of capturing time-series relationships at different scales. It combines encoder and decoder layers to capture temporal dependencies and temporal-attention layers to capture long-range dependencies across the encoded degradation trends. Temporal attention is a self-attention mechanism that enables the model to consider the importance of each time step degradation in the context of the entire degradation profile of the given health parameter. Performance data from multiple two-spool turbofan engines is employed to train and test the method. The test results show promising forecasting ability of the proposed method multiple flight cycles into the future. By leveraging the insights provided by the method, maintenance events and activities can be scheduled in a proactive manner. Future work is to extend the method to estimate remaining useful life.

Nomenclature

AI	artificial intelligence
ARIMA	auto-regressive moving average
CMAPSS	commercial modular aero-propulsion system simulation
CNN	convolutional neural network
d_K	gradient scaling factor
GRN	gated residual network
HPC	high-pressure compressor
HPT	high-pressure turbine
K	key matrix
LPC	low-pressure compressor
LPT	low-pressure turbine
LSTM	long short-term memory
m_n	slope of a normal degradation
NLP	natural language processing
PSO	particle swarm optimisation

Q	query matrix
OEM	original equipment manufacturer
RMSE	root mean squared error
RNN	recurrent neural network
RUL	remaining useful life
t	time
t_E	excitation energy
T	transpose
TFT	Temporal Fusion Transformer
V	value matrix
δa	abnormal degradation
δn	normal degradation
δw	initial wear
ε	process noise

1.0 Introduction

The modern aviation and power generation industries heavily rely on the efficiency, reliability and safety of gas turbine engines. These complex mechanical systems play a pivotal role in a wide range of applications, from propelling aircraft to generating electricity. Ensuring their optimal performance, minimising downtime and preventing catastrophic failures are critical objectives to achieve operational excellence and cost-effectiveness. In this pursuit, the field of gas turbine diagnostics and prognostics has emerged as a promising avenue for enhancing maintenance practices and extending the operational lifespan of these vital machines.

Unlike gas turbine diagnostics, the scope of gas turbine prognostics literature is comparatively more limited. The available prognostic methods can be categorised into physics-based [1, 2], data-driven [3, 4] and hybrid-physics-data-driven [5, 6]. A review of model-based and data-driven gas turbine prognostic methods was provided by Tahan et al. [7]. According to Byington et al. [8], early methods often relied on expert knowledge, first principles, empirical relationships and AI methods. These methods were mainly used to estimate a remaining useful life (RUL) based on degradation signatures and observed trends in engine sensor data. DePold et al. [9] used knowledge-based expert systems and neural networks for a combined diagnostic-prognostic system. Roemer and Kacprzynski [10] evaluated the value of prognostics in risk assessment for decision-making and discussed the significance of data fusion in gas turbine prognostics. Marinai et al. [11] applied ARIMA (auto-regression integrated moving average) to gas turbine degradation prediction, with the help of physics-based mathematical degradation models.

Following the publication of the NASA's Commercial Modular Aero-Propulsion System Simulation (CMAPSS) dataset [12] and the enhanced version N-CMAPSS dataset [13], recent advancements in gas turbine prognostics have primarily focussed on deep learning methods [14]. Notably, techniques such as convolutional neural networks (CNNs) [15, 16], long short-term memory (LSTM) [17, 18], and combined CNN-LSTM [19, 20] were employed. Chao et al. [6] proposed a physics-informed deep LSTM approach for RUL prediction in fleets of turbofan engines. The hybrid framework uses physics-based models to estimate hidden parameters related to system health, which are then fused with sensor data to create an LSTM based prognostics model. LSTM and other recurrent neural networks (RNNs) have gained prominence in sequential data processing tasks due to their capability to model temporal dependencies and sequential patterns in sensor data. These methods are powerful at capturing short-term dynamics within sequences but might encounter difficulties when dealing with long-term dependencies and complex interactions between different variables.

While deep learning methods have shown promising performance in predicting the RUL of aircraft engines, applications using Transformer models have been limited. More recently, Xia et al. [21], Hu et al. [22], Fan et al. [23] and Zhang et al. [24] applied Transformer nets to predict the RUL of turbofan engines through supervised regression learning. They utilised the CMAPSS dataset to train and validate their proposed methods. The problem was treated as a time-series regression task, using sets of

gas path measurements as input to predict the RUL of different engine components as the target variable. It has been shown that transformer models outperform CNN and LSTM.

There are two general categories of prognostics in machines health management. The first, and more common, type is RUL prediction based on the current and past degradation profiles. RUL refers to the time left to approach the predefined acceptable operating margins of the machine or its components. For gas turbines, indicators such as exhaust gas temperature (EGT) margins and gas path components' stall margins are commonly used to define RUL [12]. These margins can be defined by users or original equipment manufacturers (OEMs). As highlighted above, previous studies on gas turbine prognostics focussed solely on RUL estimation. However, in harsh operating conditions, it is crucial to predict the likelihood that the machine/engine will keep working without significant performance loss or potential damage until some future scheduled inspection [25]. This requires effective degradation tracking and forecasting system.

Fouling contributes to 70% to 85% of performance deterioration in gas turbines by impacting the efficiency of components and reducing the airflow across the gas path [26]. Fouling induced performance degradation is temporary and can be restored through effective cleaning/washing. There exist two types of washing in gas turbine compressors: online/in-service washing and offline/out-of-service washing. Online washing involves injecting water or water-based detergent solution into the compressor(s) during operation. It is relatively gentle and helps prevent the buildup of foulant deposits. Offline washing (also known as crank washing) is conducted after the gas turbine is turned off and cooled. It involves a more thorough cleaning process, often utilising strong detergents and hand cleaning, to effectively remove accumulated deposits. The downtime associated with more frequent offline washing can result in decreased overall availability and productivity of the gas turbine system, potentially leading to financial losses or disruptions in service delivery.

There are various recommendations in the literature regarding the frequency of the two washing mechanisms, with a focus on cost-benefit analysis [27–29]. Stalder [30] suggested online washing with water every day to every three days, followed by detergent cleaning at least once a week, depending on the type of foulant. Through a comprehensive analysis on data from six gas turbines, daily compressor online washing coupled with regular interval offline washing was suggested for gas turbines operating at high fouling rate [31]. It was recommended to carry out at least four crank washings per year to remove accumulated deposits downstream [30, 32].

However, strict adherence to offline washing recommendations optimised solely through cost-benefit analysis is challenging. Firstly, there is no consensus in the literature on washing frequency due to differences in operating environments between land-based and aero-engines. Even among land-based engines or aero-engines, exposure to fouling agents can vary based on engine configuration, specific site location, and surrounding environment, impacting the optimal frequency and method of compressor washing for each engine type. Secondly, predefined schedules may not account for sudden environmental changes, leading to unforeseen performance declines and downtimes. Therefore, offline washing schedules should be based on actual degradation rates, determined through diagnostic and prognostic methods. This approach could offer the most effective approach for all gas turbine applications and conditions [33].

There have been limited recent studies focusing on tracking fouling trends using deep learning for compressor offline washing optimisation. Chen et al. [34] employed LSTM for fouling trend tracking in a single-shaft industrial gas turbine compressor. They integrated the fouling prediction model, the gas turbine performance model and an economic model of compressor washing for the optimisation. A particle swarm optimisation (PSO) algorithm is employed to determine the optimal offline washing schedule with variable intervals. This integration led to better financial outcomes compared to other cost/benefit-based optimisation techniques for the washing scheduling. A similar approach has also been presented in Ref. [35].

The primary aim of this paper is to develop a method for forecasting compressor fouling trends in gas turbines. It proposes a novel Temporal Fusion Transformer (TFT) algorithm capable of tracking and forecasting flow capacity trends due to fouling in aircraft engines. It then illustrates how predictions of flow capacity degradation trends can inform the scheduling of offline compressor washing in advance.

Performance data from multiple high-bypass ratio-geared turbofan engine configurations for short-range applications are considered to demonstrate and validate the effectiveness of the proposed method. The main contributions are:

1. Use of transformer models for fouling trend prediction and forecasting: The proposed TFT-based compressor fouling trend tracking and forecasting method uses historical time-series data to estimate the current fouling level in terms of flow capacity loss and project the trend into the future. The method can also be applied to predict component efficiency trajectories, providing a comprehensive approach to prognose other failure causes. This approach has not yet been implemented in the field of gas turbine engine health management systems in this way.

2. Compressor offline washing scheduling based on flow capacity predictions: instead of strictly following OEM recommended compressor offline washing intervals, using fouling trend predictions to suggest upcoming crank washing events in advance could be more reliable and economical. There are quite limited studies available in demonstrating fouling trend predictions for compressor washing optimisation.

3. Robustness to handle engine-to-engine performance variations: it was also verified that the method proposed is advantageous in handling a considerable disparity between the training and test datasets, which is difficult for most of the traditional data-driven methods. This robustness is important to accommodate engine-to-engine degradation profile differences.

2.0 Method

2.1 Temporal fusion transformer

The Transformer model is a special kind of neural network structure introduced by Vaswani et al. in 2017 [36]. It is a breakthrough in natural language processing (NLP) and quickly evolved as one of the most powerful methods for sequence-to-sequence problems. The structure is composed of sets of encoder and decoder blocks, each with multi-head-self-attention and feed-forward layers. The self-attention scheme involved in the structure assigns weights to different parts of the input data based on their significance to the output. The feed-forward layers process the outputs of the self-attention layers and perform the final prediction. Different advancements and modification have been made to the original Transformer architecture over time including Transformer-XL, Bidirectional Encoder Representations from Transformers (BERT), Performer BERT, and Generative Pre-trained Transformer (GPT) [37].

Temporal Fusion Transformers, on the other hand, are specifically designed for time-series prediction and forecasting [38]. They extend the Transformer architecture by incorporating autoregressive modelling, exogenous inputs and a gating mechanism to control the flow of information between the different components. In addition, TFTs introduce a new type of attention mechanism called Temporal Attention, which enables the model to attend to both past and future inputs when making predictions. The core idea behind TFTs is to model the temporal dependencies in time-series data by using a self-attention mechanism. The self-attention mechanism allows the model to attend to different parts of the input sequence and weight them differently depending on their relevance to the output prediction. TFTs have shown promising results on a variety of time-series forecasting tasks, including predicting energy consumption, traffic volume and weather prediction [38, 39].

2.2 Gas turbine prognostics using TFT

The TFT model aims to predict the current performance of a gas turbine and forecast the degradation progress in the future time steps, given historical measurements up to the current time. The methodology involved a series of steps to prepare the data, design the model architecture and train it for accurate predictions.
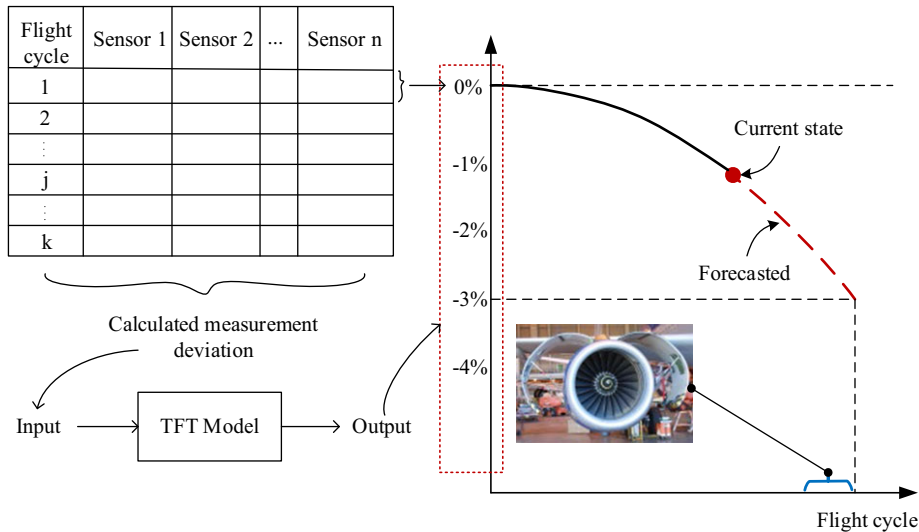
Figure 1. *Input/output structure of the TFT model.*

In the first place, the time-series performance data needs to be structured in a format that can be fed into the TFT model. This typically involves converting each observation into a fixed-size vector representation that the TFT model can process. This is done by combining temporal positional embeddings and feature embeddings. Different ranges of vector representation specify the dimensionality of the input features (i.e. the number of selected gas path measurements) and the length of the input sequence or window size. The window size indicates how many previous observations in the time-series are used to predict flow capacity and efficiency deviation in the next timestep. Input-output pairs are then formed for the training. The inputs are the embedded segments of measured values, while the outputs are degradation magnitudes to be predicted. Figure 1 illustrates the structure of input-output performance data to the TFT model before the embedding. It is good to note here that the current state can be estimated by a diagnostic system. When the estimated degradation by the diagnostic system reaches a predefined threshold to start the forecast, the prognostic model will be activated to forecast the progress.

The structure of the TFT model for degradation tracking and forecasting was determined through experimentation. Various numbers of layers for both the encoder and decoder were examined. Each encoder block primarily comprises input processing layers, static covariate and LSTM-encoder layers, and Add & Normalisation layer. Each decoder block consists of multi-headed feed-forward layers, Add & Normalisation layers, dropout layer and Dense layers. The encoders process the historical gas path measurements, capturing temporal dependencies using LSTM networks and relevant patterns using multi-headed self-attention layers. The decoder generates degradation forecasts based on the encoded input representations. It uses the attention mechanism to focus on relevant parts of the encoded information and combines it with its own processing to make forecasts. For the underlying mathematics and detail descriptions of the encoder and decoder layers, interested readers are encouraged to refer to Ref. [36]. Figure 2 shows a simplified layout of the TFT architecture applied.

2.2.1 Description to the major layers of the TFT model

- LSTM-encoder/decoder layers: The TFT structure involves sets of LSTM-based encoder and decoder layers to capture temporal relationship between past and future degradation trends in the time-series and forecast future values.
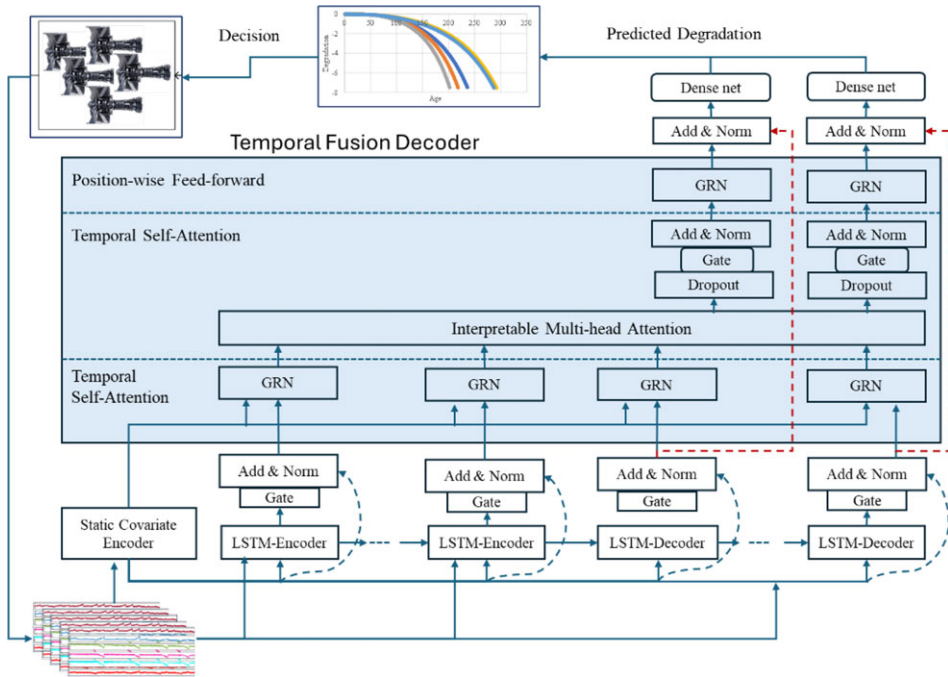
Figure 2. *Schematics of the TFT architecture used to predict gas turbine compressor fouling trends for washing scheduling, adapted from Ref. [36].*

- Static covariate encoders: These layers enable incorporating static features into the encoder and decoder layers to give some context to the temporal dynamics.

- Multi-headed self-attention layers: This module incorporates several attention layers computing in parallel to capture long term dependencies between different parts of the input sequence. Mathematically:

$$Attention\,(Q, K, V) = \text{softmax}\left(\frac{QK^{T}}{\sqrt{d_k}}\right) V \tag{1}$$

where Q, K and V are the query matrix, key matrix and value matrix, respectively, SoftMax computes the attention scores for each query-key pair, and $\sqrt{d_k}$ is a scaling factor to prevent extremely small gradients during training.

- Add & norm layers: These layers add the residual connection from the previous layer and applies normalisation. Residual connections are used to avoid the vanishing gradient problem, and normalisation stabilises the training process.

- Feed-forward layers: This part introduces non-linearity to the data and further learns useful features from it. The default feed-forward option is 'Gated Residual Network (GRN)'. There are also other alternatives including 'Gated Linear Unit (GLU)', 'Bilinear', 'ReGLU', 'Gated Enhanced Gated Linear Unit (GEGLU)', 'SwiGLU'.

- Dense nets: These are fully connected regression layers used to estimate the degradation value for each timestep. They also provide probabilistic forecasts by estimating quantiles of the predictive distribution. This allows the model to capture prediction uncertainty and provide interval forecasts.

Figure 3. *Schematic diagram of a two-spool turbofan engine.*

2.3 Using degradation forecasts to schedule maintenance events

The frequency of offline compressor washing depends on several factors such as engine type, weather condition, fouling rate, etc. The best option to handle these factors could be tracing fouling effects using diagnostic and prognostic tools and associating it with the washing scheduling. Some authors, like Ref. [40, 41], suggested conducting offline compressor washing when the flow capacity loss due to fouling reaches 2%–3%. As discussed in Section 2.2, the proposed TFT method can track and forecast flow capacity degradation trajectories. Hence, instead of following a fixed offline washing intervals despite the gas turbine type and the actual fouling rate, it would be more beneficial to use the TFT model to predict when the flow capacity loss due to fouling reaches 2%–3%.

Although this paper primarily focuses on compressor washing, the TFT algorithm can also predict hot section degradation profiles and utilise this information to schedule overhaul events. This task needs integrating the prognostic system with an effective diagnostic system, to identify affected components and root causes first. In a previous study, the authors of this paper proposed a hybrid physics-based and data-driven diagnostic system that can effectively detect and isolate gas turbine faults to the component level [42, 43]. The current method can then be coupled with the diagnostic system to have an integrated diagnostic-prognostic framework. Once the root cause is identified and the current severity is estimated by the diagnostic system, the prognostic model associated with the affected component can forecast the future progress of the degradation and estimate the overhaul time based on economic considerations.

2.4 Case study on turbofan engines

To demonstrate the performance of the method proposed, a synthetic dataset was generated from a two-spool turbofan engine through an in-house dynamic model, EVA. EVA is a physics-based performance code employing Gibbs free energy minimisation for combustion modeling and comprehensive component modeling. It integrates first and second-order physical effects, including non-linear off-design behaviour through component maps. Its solver, a gradient-based Newton-Raphson approach, ensures robustness. EVA's object-oriented design facilitates flexibility and customisation for various gas turbine configurations, adhering to international standards. Detail information about the software can be found in Refs [44, 45]. Figure 3 shows a schematic representation of the case study engine along with the corresponding station numbers. The method used to generate degradation trajectories followed the well-established N-CMAPSS dataset generation procedure [13], and the main steps are highlighted as follows:

1. Define input conditions: The simulation was performed at a cruise flight segment, and a snapshot was considered for each flight cycle simulation. For every engine considered in the simulation, a similar Mach number (M = 0.7), altitude (Alt = 20 Kft), and flight length was considered.

2. Define degradation trajectories: The health condition of the major rotary components was represented by efficiency and flow capacity parameters. The evolution of the degradation trajectories begins with some level of initial wear, to account the production and assembly tolerance among fleets of engines. It was modelled by adding normally distributed random values ranging 0 to 1% [12]. Following the establishment of initial wear, the normal degradation effect is added into the trajectory. This phase simulates the gradual decline in engine performance over time due to usage, often considered as normal and is modelled as linear decreasing trend. Each sub-component is assumed to have a maximum excitation energy threshold. This threshold represents the limit beyond which the sub-component transitions to an abnormal state. The distribution of maximum excitation energy threshold values is modeled as Gaussian to account for material property variability in each unit. The abnormal degradation onset time due to a fault is defined as the point in time when the total excitation energy that a component has experienced from the initial time ($t = 0$) to a specific time (t_E) exceeds the maximum excitation energy threshold. In the current work, short flight lengths (up to three hours) with low and medium operation intensity were considered for the case study engines. Abnormal degradation patterns are then introduced to the trajectory, based on the exponential decay model provided [by] Saxena and Goebel [12]. These patterns simulate typical degradation events induced by rapid or abrupt failure conditions. These anomalies can accelerate degradation, leading to rapid declines in engine health. The degradation model used to compute these trajectories is given in Equations (2) and (3). The simulation continues up to a point, signifies a critical threshold beyond which the engine can no longer operate effectively.

$$\delta_n(t) = m_n(t) + \delta_{iw} \quad \forall t < t_E \tag{2}$$

$$\delta_a(t) = 1 - exp\left[a(t - t_E)^b\right] + \delta_n(t_E) + \varepsilon \tag{3}$$

where δ_{iw} is initial wear (where $\delta_{iw} \in [0\%, 1\%]$), $m_n = -0.001$ is slope of the normal degradation $\delta_n(t)$, t_E is the flight cycle where the excitation energy that the particular component has experienced exceeds the maximum, ε is process noise, a = U(0.001, 0.003), b = U(1.4, 1.6) and ε = N(0, 0.001) for efficiency and ε = N(0, 0.003) for flow capacity. For a given engine, a and b are constant. The maximum tolerated degradation considered for all flow capacities was $\pm 7.5\%$, while for all efficiencies was -4%.

3. Generate sensor data: We impose degradation trajectories obtained from step 2 to the engine performance model and simulate measurements for a fleet of 30 engines, each with a different degradation mode. Intermediate pressure compressor (IPC) exit total temperature (T25), IPC exit total pressure (P25), high pressure compressor (HPC) exit total temperature (T3), HPC exit total pressure (P3), high pressure turbine (HPT) exit total temperature (P43), HPT exit total pressure (P43), low pressure turbine (LPT) exit total temperature (T5), low pressure shaft speed and high pressure shaft speed sensors were included in the model.

4. Add senser noise: After generating the required data, a white Gaussian noise was added to accommodate sensor non-repeatability. Signal-to-noise ratio of 60dB was considered as maximum noise level for each sensor involved.

5. After forming the input-output data pairs, the dataset was divided into training and test sets. Each dataset was normalised to ensure better model performance and convergence. Data from 25 engines were utilised for training and from 5 engines for test. The trained TFT model was then used to forecast degradation trajectories into the future timesteps.

Table 1. *Hyperparameters of the proposed TFT model*

Hyperparameter	Value
Feature dimension	12
Window size	32
Number of hidden units	68
Number of recurrent layers	4
Number of attention heads	4
Dimensions of the feedforward network	32
Batch size for training	64
Activation function used in the model	ReLU
Optimisation algorithm	Adam
Learning rate	0.001
Number of epochs	20
Dropout rate	0.12

3.0 Results and discussion

To generate a single time sequence data, the entire degradation trajectories considered for training and testing were stacked one after the other, and the associated input measurements were treated similarly. After this, the data was prepared (embedded and encoded) according to the specific input format requirements of the TFT model. Subsequently, the data was divided into training, testing and forecast subsets. The model was trained using the training set to determine its parameters. The test dataset, which chronologically follows the training data, was then used to assess the generalisation performance of the method and fine-tune the hyperparameters through an iterative process. After training and testing the model, it was used to generate forecasts for the future. In all three cases, the predicted values were compared to the actual values to calculate the RMSE of the predictions.

During the model's hyperparameter tuning process, the training data was partitioned into varying batch sizes, ranging from 32 to 160. Several iterations were conducted to determine the optimal hyperparameters. This involved evaluating different feedforward dimensions (ranging from 8 to 256), incrementally adjusting dropout probabilities from 0.1 to 0.5, varying batch sizes from 32 to 256, evaluating learning rates at 0.1, 0.01, 0.001 and 0.0001, and considering the number of attention, encoder and decoder heads, ranging from 1 to 8 for each component. Ultimately, the model with the best performance was selected, a summary of the hyperparameters of the selected model is shown in Table 1.

In this paper, only results for the IPC flow capacity prognostics are presented aiming to demonstrate how the forecast could be incorporated into compressor washing scheduling. However, the method can also be used to predict the other health parameters in the gas path in a similar way. Figure 4 displays the test and forecast results obtained. Flow capacity patterns from 25 engines and the associated simulated sensor data were used for training. Degradation data from engine 26 to 29 (up to end-of-life), and the degradation data from engine 30 up to a random degradation level were used as a test dataset. The forecasting task was initiated randomly when the flow capacity deviation in the degradation trajectory of engine 30 reached -1.64%. RMSE values of 0.1128, 0.1121, 0.0978, 0.1345, 0.1739 and 0.273 were obtained for 10, 20, 30, 40, 50 and 60 flight cycles into the future, respectively.

3.1 Effect of the forecast starting time

To assess the method's sensitivity to the initial degradation level for trajectory forecasting, three different scenarios were examined: Scenario 1 (starting at -1%), Scenario 2 (starting at -1.5%) and Scenario 3 (starting at -2%). Forecasting continued until the degradation level reached -3% from each reference point. The resulting RMSE values were 0.0963, 0.0973 and 0.0904, respectively. As shown in Fig. 5, in
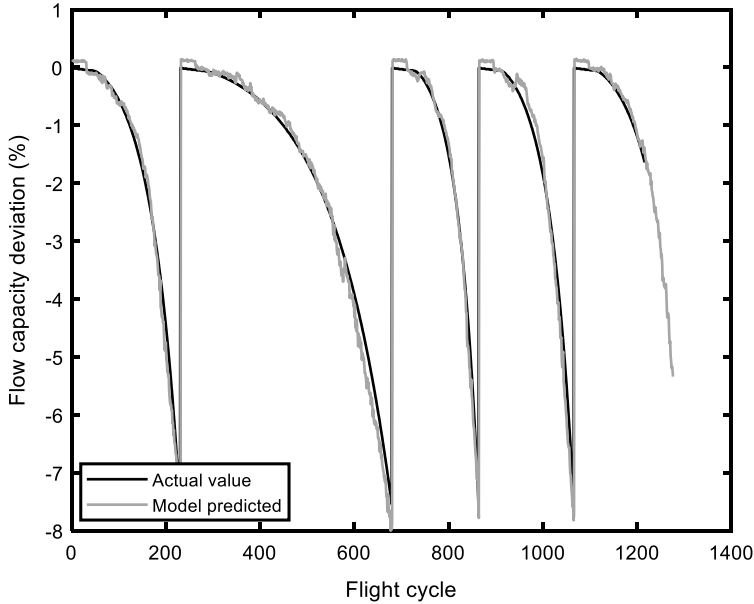
Figure 4. *Displays the test and forecast results of the TFT model for IPC flow capacity, showing the forecast from a random initial degradation level (−1.64) up to 60 future flight cycles.*
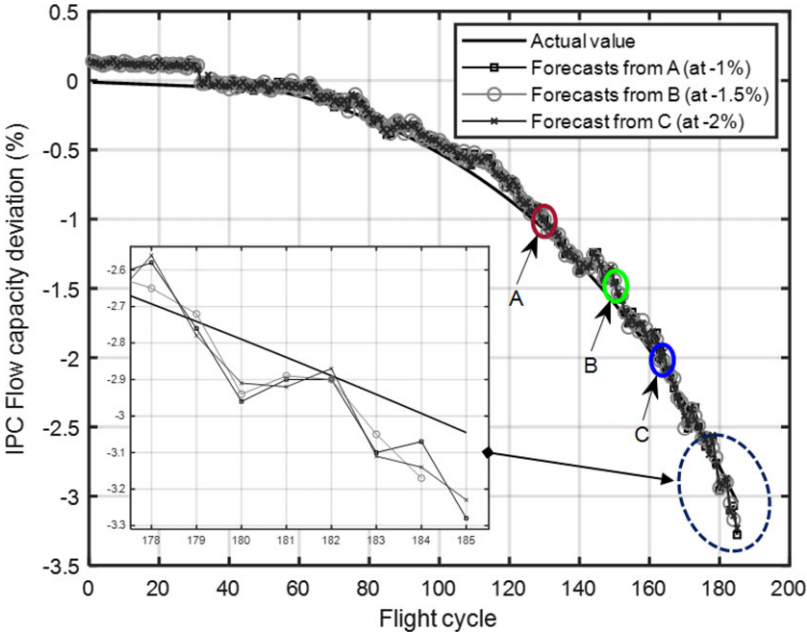

Figure 5. *IPC flow capacity loss forecasts from different initial points: −1%, −1.5 and −2%. From each initial point the forecast continued up until −3% flow capacity loss.*
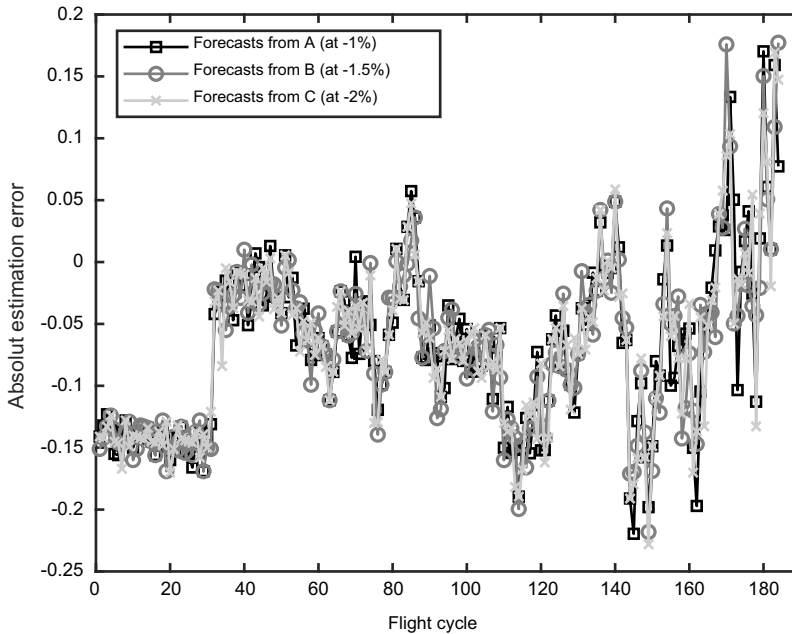
Figure 6. *Test and forecast errors of the trajectory of engine 30 flow capacity deviation.*

all three scenarios, the target -3% flow capacity was predicted two flight cycles earlier than expected. However, since offline washing is typically recommended within the range of 2%–3% flow capacity loss, the impact of these observed errors on the offline washing scheduling is negligible. Early prediction is preferrable, as late prediction could result in an in-service failure, which could have safety, economic and maintenance efficiency implications. Setting the washing time when the estimated flow capacity loss reaches -2.5% would provide a high-confidence margin in case of any unexpected prediction outliers.

The absolute error for engine 30's trajectory, including the test and forecasted horizons, is presented in Fig. 6. In the initial and linear degradation phases of the IPC, the prediction errors were large. This was mainly caused by measurement noise during normal operation. As the degradation gradually developed, the TFT model improved its ability to track the target degradation profile. Similarly, as the length of the forecasting time step into the future increased, deviations between the forecasted and actual degradation increased, as expected, but they remained within an acceptable range.

3.2 *Effect of the training and test dataset ratio*

In machine learning, it is common practice to split the dataset into training and test sets, with around 70%–80% of the data allocated for training and the remaining 20%–30% for testing. The model learns from the training dataset to make predictions and is evaluated on the unseen test dataset to assess its performance. Additionally, during the training phase, it's typical to further split the training dataset into training and validation sets. The validation set is used for model tuning and hyperparameter optimisation. In this study, we investigated the impact of different test-to-training data ratios, including 60/40, 70/30 and 90/10. In all three cases, forecasting commenced when the flow capacity deviation of the last trajectory reached -1% and continued until it reached -3% flow capacity deviation.

The prediction accuracy of both the test and forecast data was evaluated in terms of root mean square error (RMSE) and summarised in Table 2. For visualisation purposes, the comparison between the target degradation trajectories vs. the model predicted values for these three cases is also displayed in Fig. 7. The test results showed that for the 60/40 ratio, an RMSE of 0.4833 was achieved, while the 70/30 ratio

Table 2. *Performance of the TFT model at different training and test dataset ratio*

Case	Training/Test ration	Forecast starting	Forecast horizon	Test RMSE	Forecast RMSE
1	60/40 ratio (or 18/12 engines)			0.4833	Error up to −3% forecast = 0.4263 Error up to −2.5% forecast = 0.4167
2	70/30 ratio (or 21/9 engines)	1% flow capacity loss	144 flight cycles into the future (i.e., up to −3% flow capacity loss)	0.4749	Error up to −3% forecast = 0.3369 Error up to −2.5% forecast = 0.3289
3	90/10 ratio (or 27/3 engines)			0.4622	Error up to −3% forecast = 0.2406 Error up to −2.5% forecast = 0.2329

resulted in an RMSE of 0.4749, and the 90/10 ratio resulted in an RMSE of 0.4622. When compared to the test RMSE of 0.4396 obtained for the 80/20 ratio (as described in Section 3.1), the RMSE values for these three scenarios were slightly lower. For the 60/40 ratio, the RMSE value of the forecast up to −3% flow capacity loss is 0.4263. This implies that the TFT model prediction suggests washing when the actual degradation falls between −2.5% and −3%, in agreement with the recommended range of 2%–3% flow capacity deviation. The forecasting error up to −2.5% flow loss is an RMSE of 0.4167. This means that if washing is planned based on a predicted value of −2.5% flow capacity deviation, the washing will be performed prior to the actual flow capacity loss reaching −2.5% due to fouling, but still within the recommended range of 2%–3% flow capacity loss. Similarly, for the second case (the 70/30 ratio), the RMSE of the forecast up to the −3% flow capacity loss is 0.3369, and the forecasting error up to −2.5% flow loss is an RMSE of 0.3289. For the last case, the RMSE of the forecast up to the −3% flow capacity loss is 0.2406, and the forecasting error up to −2.5% flow loss is an RMSE of 0.2329.

In general, based on the TFT model prediction, scheduling offline washing within estimated −2.5% to −3% flow capacity deviation ensures that compressor washing will be performed within the recommended 2%–3% flow capacity loss. Earlier scheduling is a safer approach compared to having washing events occurring later than the expected −3% flow capacity change.

3.3 Effect of the abnormal degradation starting time

The time to exceed the maximum excitation energy is strongly correlated with the inlet temperature [46]. Harsh flight conditions and long flights often result in a relatively rapid shift from a linear to an abnormal degradation trend. To assess the performance of the method on degradation profiles distinct from those employed during training and testing, due to variations in the timing of maximum excitation energy, four different cases were considered: trajectories with time to maximum excitation energy $(t_E) = 20, 40, 60$ and 80 flight cycles. The degradation trajectory constants 'a' and 'b,' as well as the process noise 'ϵ,' were randomly chosen as discussed in Section 2.4. In both the training and test trajectories, a constant t_E ($t_E = 100$ flight cycles) was utilised. The results of the test are presented in Fig. 8, where trajectories B, D, F and H represent the actual degradation trajectories for the four case studies, and trajectories A, C, E and G represent the corresponding predicted trajectories.

The RMSE of the forecast was approximately 0.26. If a −2% flow capacity deviation is considered as the lower threshold for triggering offline washing, there is a possibility that the model may predict the need for washing 10 to 15 flights earlier than the expected time. Based on the RMSE value and as
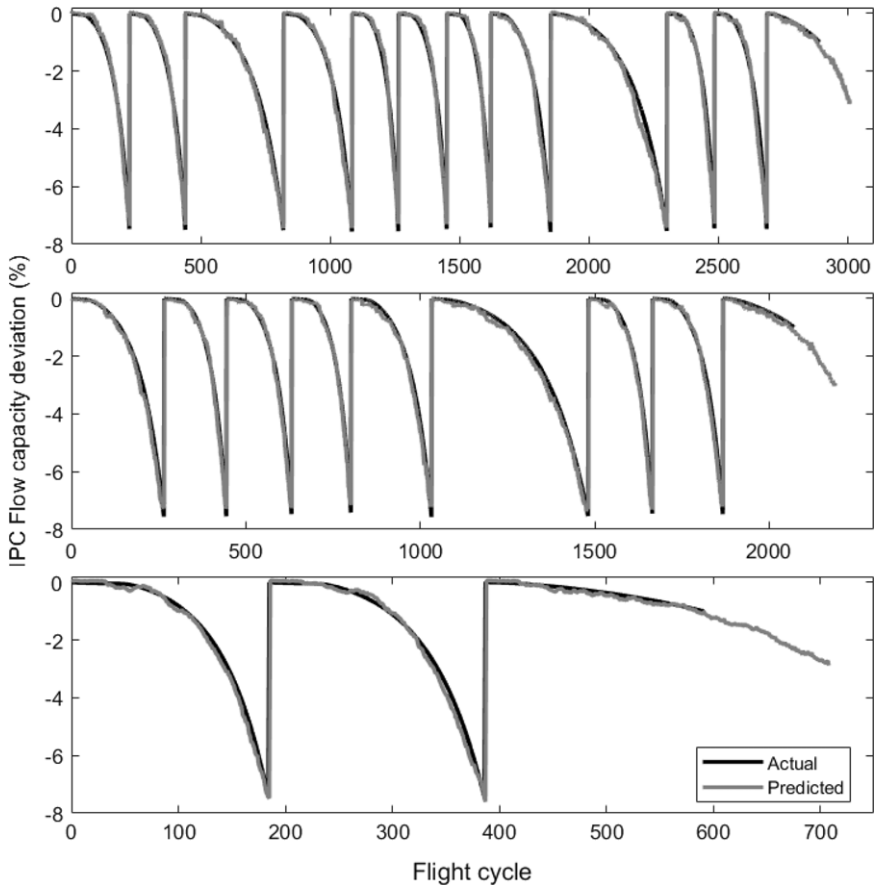
Figure 7. *Test and forecast results are presented for three training/test data ratios. The top subplot corresponds to the 40% test dataset, the middle subplot to the 30% test dataset and the bottom subplot to the 10% test dataset.*

shown in the zoomed in figure (Fig. 8(b)), the flow capacity changes observed between 10 to 15 flight cycles are generally less than −0.26% within the target −2% to −3% region. Therefore, scheduling the offline washing event based on a threshold of −2.3% to −3% of the model prediction would be safe.

4.0 Conclusions

This paper has introduced a novel prognostic method for forecasting gas turbine degradation based on time-series measurements. The Temporal Fusion Transformer model, which combines encoder and decoder layers with temporal-attention mechanisms, has demonstrated its ability to accurately predict degradation trajectories of several flight cycles into the future. This allows gas turbine operators proactively schedule maintenance events, optimise compressor washing frequency and minimise unexpected downtimes.

The study demonstrated the effectiveness of the proposed method by tracking compressor degradation using flow capacity as the primary parameter. Flow capacity changes are indicative of compressor fouling deposit progression, and these changes inform washing interval scheduling. The method can effectively forecast offline washing events in advance when fouling deposits reach detectable levels. It is essential to note that this method can also be applied to predict component efficiency trajectories,
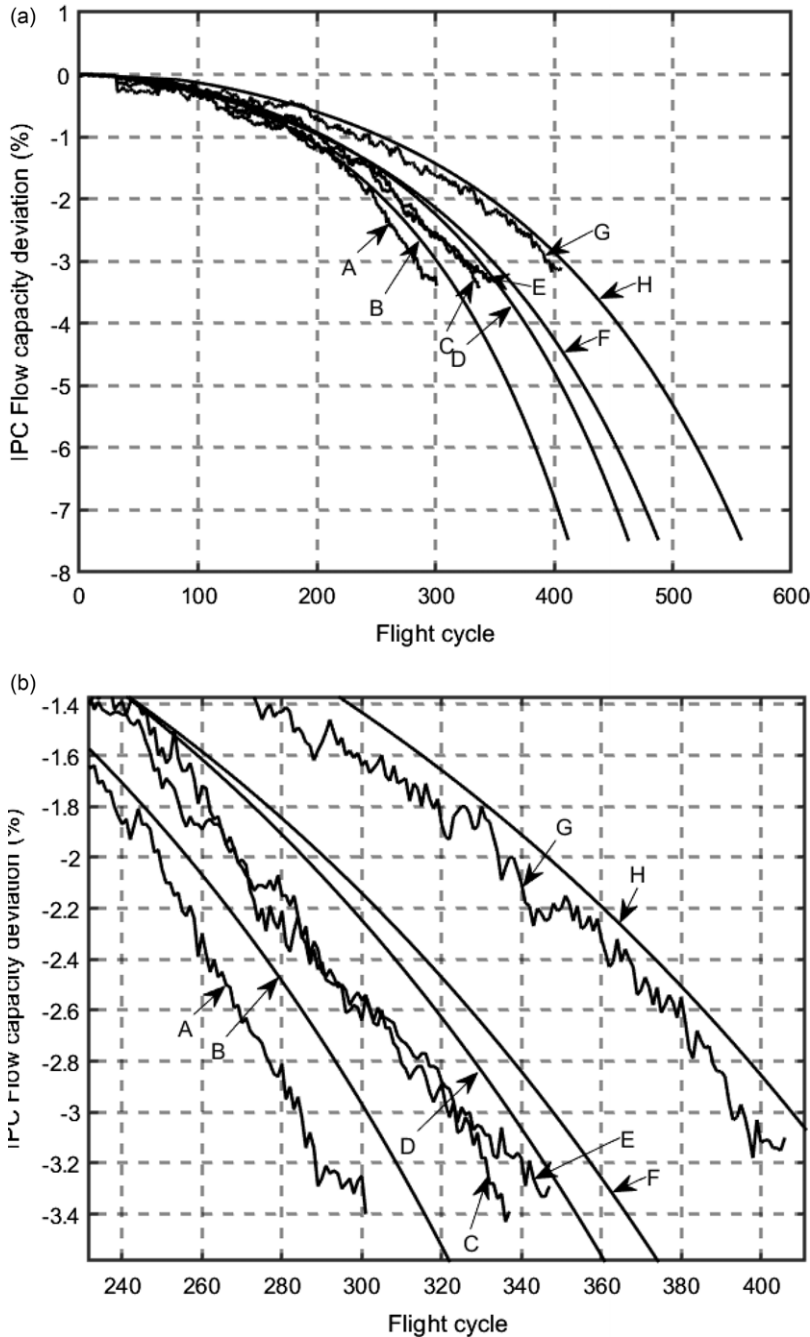
Figure 8. *Flow capacity forecasting results for different degradation trajectories: (a) forecasted vs. target degradation trajectories and (b) zoomed in view of the forecast between −1.4% and −3.4.*

providing a comprehensive approach to prognose all types of gas path faults. Integrating the proposed prognostic method with a diagnostic method enables a holistic approach to assess engine component(s) health and predict maintenance intervals.

Future work in this area may include extending the method to estimate the remaining useful life of gas turbine components, conducting further analysis of forecasting accuracy while increasing the number of

degradation trajectories considered for training, exploring the integration of economic analysis into the washing optimisation process and applying this methodology to a real engine to assess its performance in real-world scenarios.

Competing interests. The authors declare that they have no competing interests.

References

[1] Li Y.G. and Nilkitsaranont P. Gas turbine performance prognostic for condition-based maintenance. *Appl. Energy*, 2009, **86**, (10), pp 2152–2161. doi: 10.1016/j.apenergy.2009.02.011

[2] Alozie O., Li Y.-G., Wu X., Shong X. and Ren W. An adaptive model-based framework for prognostics of gas path faults in aircraft gas turbine engines. *Int. J. Prognost. Health Manag.*, 2019, **10**, (2). doi: 10.36001/ijphm.2019.v10i2.2728

[3] Losi E., Venturini M. and Manservigi L. Gas turbine health state prognostics by means of Bayesian hierarchical models. *Proceedings of the ASME Turbo Expo 2019: Turbomachinery Technical Conference and Exposition. Volume 9: Oil and Gas Applications; Supercritical CO2 Power Cycles*, Wind Energy. Phoenix, Arizona, USA. June 17–21, 2019, p V009T27A003. doi: 10.1115/GT2019-90054

[4] Zaccaria V., Fentaye A.D. and Kyprianidis K. Bayesian information fusion for gas turbines diagnostics and prognostics. *ASME Turbo Expo 2023: Turbomachinery Technical Conference and Exposition, 2023, vol. Volume 4: Controls, Diagnostics, and Instrumentation*, p V004T05A021, doi: 10.1115/gt2023-103171

[5] Kordestani M., Mousavi M., Chaibakhsh A., Orchard M.E., Khorasani K. and Saif M. A new compressor failure prognostic method using nonlinear observers and a Bayesian algorithm for heavy-duty gas turbines. *IEEE Sens. J.*, 2023, **23**, (4), pp 3889–3900. doi: 10.1109/JSEN.2022.3233585

[6] Arias Chao M., Kulkarni C., Goebel K. and Fink O. Fusing physics-based and deep learning models for prognostics. *Reliab. Eng. Syst. Saf.*, 2022, **217**, p 107961. doi: 10.1016/j.ress.2021.107961

[7] Tahan M., Tsoutsanis E., Muhammad M. and Abdul Karim Z.A. Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: a review. *Appl. Energy*, 2017, **198**, pp 122–144. doi: 10.1016/j.apenergy.2017.04.048

[8] Byington C.S., Roemer M.J. and Galie T. Prognostic enhancements to diagnostic systems for improved condition-based maintenance [military aircraft]. *Proceedings, IEEE Aerospace Conference*, Big Sky, MT, USA, 9–16 March 2002, 2002, 6, p 6. doi: 10.1109/AERO.2002.1036120

[9] DePold H.R. and Gass F.D. The application of expert systems and neural networks to gas turbine prognostics and diagnostics. *J. Eng. Gas Turbines Power*, 1999, **121**, (4), pp 607–612. doi: 10.1115/1.2818515

[10] Roemer M.J. and Kacprzynski G.J. Advanced diagnostics and prognostics for gas turbine engine risk assessment. *2000 IEEE Aerospace Conference. Proceedings (Cat. No.00TH8484)*, Big Sky, MT, USA, 25–25 March 2000, 2000, 6, pp 345–353. doi: 10.1109/AERO.2000.877909

[11] Marinai L., Singh R., Curnock B. and Probert D. Detection and prediction of the performance deterioration of a turbofan engine. *Proceedings of the International Gas Turbine Congress*, Tokyo, 2–7 November 2003, Gas Turbine Society of Japan (GTSJ), 2003, pp. 2–7.

[12] Saxena A., Goebel K., Simon D. and Eklund N. Damage propagation modeling for aircraft engine run-to-failure simulation. *2008 International Conference on Prognostics and Health Management*, Denver, CO, USA, IEEE, 2008, pp 1–9. doi: 10.1109/PHM.2008.4711414

[13] Arias Chao M., Kulkarni C., Goebel K. and Fink O. Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics. *Data*, 2021, **6**, (1), p 5. https://www.mdpi.com/2306-5729/6/1/5

[14] Ramasso E. and Saxena A. Performance benchmarking and analysis of prognostic methods for CMAPSS datasets. *Int. J. Prognost. Health Manag.*, 2014, **5**, (2), pp 1–15. doi: 10.36001/ijphm.2014.v5i2.2236

[15] Li X., Ding Q. and Sun J.-Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab. Eng. Syst. Saf.*, 2018, **172**, pp 1–11. doi: 10.1016/j.ress.2017.11.021

[16] Muneer A., Taib S.M., Fati S.M. and Alhussian H. Deep-learning based prognosis approach for remaining useful life prediction of turbofan engine. *Symmetry*, 2021, **13**, (10), p 1861. https://www.mdpi.com/2073-8994/13/10/1861

[17] Deng K., et al. A remaining useful life prediction method with long-short term feature processing for aircraft engines. *Appl. Soft Comput.*, 2020, **93**, p 106344. doi: 10.1016/j.asoc.2020.106344

[18] Lin L., Liu J., Guo H., Lv Y. and Tong C. Sample adaptive aero-engine gas-path performance prognostic model modeling method. *Knowl.-Based Syst.*, 2021, **224**, p 107072. doi: 10.1016/j.knosys.2021.107072

[19] Al-Dulaimi A., Zabihi S., Asif A. and Mohammadi A. A multimodal and hybrid deep neural network model for remaining useful life estimation. *Comput. Ind.*, 2019, **108**, pp 186–196. doi: 10.1016/j.compind.2019.02.004

[20] Kong Z., Cui Y., Xia Z. and Lv H. Convolution and long short-term memory hybrid deep neural networks for remaining useful life prognostics. *Appl. Sci.*, 2019, **9**, (19), p 4156. doi: https://www.mdpi.com/2076-3417/9/19/4156

[21] Xia J., Feng Y., Teng D., Chen J. and Song Z. Distance self-attention network method for remaining useful life estimation of aeroengine with parallel computing. *Reliab. Eng. Syst. Saf.*, 2022, **225**, p 108636. doi: 10.1016/j.ress.2022.108636

[22] Hu Q., Zhao Y. and Ren L. Novel transformer-based fusion models for aero-engine remaining useful life estimation. *IEEE Access*, 2023, **11**, pp 52668–52685. doi: 10.1109/ACCESS.2023.3277730

[23] Fan Z., Li W. and Chang K.-C. A two-stage attention-based hierarchical transformer for turbofan engine remaining useful life prediction. *Sensors*, 2024, **24**, (3), p 824. https://www.mdpi.com/1424-8220/24/3/824

[24] Zhang Z., Song W. and Li Q. Dual-aspect self-attention based on transformer for remaining useful life prediction. *IEEE Trans. Instrum. Meas.*, 2022, **71**, pp 1–11. doi: 10.1109/TIM.2022.3160561

[25] Sikorska J., Hodkiewicz M. and Ma L. Prognostic modelling options for remaining useful life estimation by industry. *Mech. Syst. Sig. Process.*, 2011, **25**, (5), pp 1803–1836. doi: 10.1016/j.ymssp.2010.11.018

[26] Diakunchak I.S. Performance deterioration in industrial gas turbines. *J. Eng. Gas Turbines Power*, 1992, **114**, (2), pp 161–168. doi: 10.1115/1.2906565

[27] Aretakis N., Roumeliotis I., Doumouras G. and Mathioudakis K. Compressor washing economic analysis and optimization for power generation. *Appl. Energy*, 2012, **95**, pp 77–86. doi: 10.1016/j.apenergy.2012.02.016

[28] Musa G., Igie U., Di Lorenzo G., Alrashed M. and Navaratne R. Gas turbine compressor washing economics and optimization using genetic algorithm. *J. Eng. Gas Turbines Power*, 2022, **144**, (9). doi: 10.1115/1.4055187

[29] Hanachi H., Liu J., Ding P., Kim I.Y. and Mechefske C.K. Predictive compressor wash optimization for economic operation of gas turbine. *J. Eng. Gas Turbines Power*, 2018, **140**, (12). doi: 10.1115/1.4041168

[30] Stalder J.-P. Gas turbine compressor washing state of the art: field experiences1. *J. Eng. Gas Turbines Power*, 2000, **123**, (2), pp 363–370. doi: 10.1115/1.1361108

[31] Schneider E., Demircioglu Bussjaeger S., Franco S. and Therkorn D. Analysis of compressor on-line washing to optimize gas turbine power plant performance. *J. Eng. Gas Turbines Power*, 2010, **132**, (6). doi: 10.1115/1.4000133

[32] Kolkman H.J. Performance of gas turbine compressor cleaners. *ASME. J. Eng. Gas Turbines Power*, 1993, **115**, (3), pp 674–677. doi: 10.1115/1.2906759

[33] Rao P.N.S. and Naikan V.N.A. An optimal maintenance policy for compressor of a gas turbine power plant. *J. Eng. Gas Turbines Power*, 2008; 130, (2), p 021801. doi: 10.1115/1.2795762

[34] Chen J., Tang X., Lu J. and Zhang H. A compressor off-line washing schedule optimization method with a LSTM deep learning model predicting the fouling trend. *J. Eng. Gas Turbines Power*, 2022, **144**, (8). doi: 10.1115/1.4054748

[35] Jin Y., et al. A hybrid model of LSTM neural networks with a thermodynamic model for condition-based maintenance of compressor fouling. *Meas. Sci. Technol.*, 2021, **32**, (12), p 124007. doi: 10.1088/1361-6501/ac026f

[36] Vaswani A., et al. Attention is all you need. *Adv. N.A. Inf. Process. Syst.*, 2017, **30**. doi: 10.48550/arXiv.1706.03762

[37] Galanis N.-I., Vafiadis P., Mirzaev K.-G. and Papakostas G.A. Machine learning meets natural language processing-the story so far. *IFIP International Conference on Artificial Intelligence Applications and Innovations*, Springer, 2021, pp 673–686. doi: 10.48550/arXiv.2104.10213

[38] Lim B., Arık S.Ö., Loeff N. and Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecasting*, 2021, **37**, (4), pp 1748–1764. doi: 10.48550/arXiv.1912.09363

[39] Wu B., Wang L. and Zeng Y.-R. Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. *Energy*, 2022, **252**, p 123990. doi: 10.1016/j.energy.2022.123990

[40] Scott J. Axial compressor monitoring by measuring intake air depression. *Third Symposium on Gas Turbine Operations and Maintenance, National Research Council of Canada Symposium*, Ottawa, Ontario, September 17–18, National Research Council, Canada, 1979. https://catalog.lindahall.org/discovery/fulldisplay?docid=alma99830793405961&context=L&vid=01LINDAHALL_INST:LHL&lang=en&search_scope=MyInstitution&adaptor=Local%20Search%20Engine&tab=LibraryCatalog&query=creator,exact,Symposium%20on%20Gas%20Turbine%20Operation%20and%20Maintenance&facet=creator,exact,Symposium%20on%20Gas%20Turbine%20Operation%20and%20Maintenance&offset=0

[41] Diakunchak I.S. Performance improvement in industrial gas turbines. *Proceedings of the 1993 Joint Power Generation Conference: GT Papers. 1993 Joint Power Generation Conference: GT Papers*, ASME, Kansas City, USA. October 17–22, 1993, 1993, p V001T01A005. doi: 10.1115/93-JPGC-GT-5

[42] Fentaye A.D., Zaccaria V. and Kyprianidis K. Aircraft engine performance monitoring and diagnostics based on deep convolutional neural networks. *Machines*, 2021, **9**, (12), p 337. doi: 10.3390/machines9120337

[43] Fentaye A., Zaccaria V. and Kyprianidis K. Discrimination of rapid and gradual deterioration for an enhanced gas turbine life-cycle monitoring and diagnostics. *Int. J. Prognost. Health Manag.*, 2021, **12**, (3). doi: 10.36001/ijphm.2021.v12i3.2962

[44] Kyprianidis K.G., Colmenares Quintero R.F., Pascovici D.S., Ogaji S.O.T., Pilidis P. and Kalfas A.I. EVA: a tool for environmental assessment of novel propulsion cycles. *ASME Turbo Expo 2008: Power for Land, Sea, and Air, 2008, vol. Volume 2: Controls, Diagnostics and Instrumentation; Cycle Innovations; Electric Power*, pp 547–556. doi: 10.1115/gt2008-50602. [Online]. doi: 10.1115/GT2008-50602

[45] Kyprianidis K. An approach to multi-disciplinary aero engine conceptual design. *International Symposium on Air Breathing Engines*, ISABE 2017, Manchester, UK, 3–8 September 2017, Paper No. ISABE-2017-22661, 2017.

[46] Kamtsiuris A.A., Raddatz F. and Wende G. A health index framework for condition monitoring and health prediction. *PHM Soc. Eur. Conf.*, 2022, **7**, (1), pp 231–238. doi: 10.36001/phme.2022.v7i1.3324