



ARTICLE

Satellite Failure Prognosis with Cascaded Temporal Convolution and Transformer Network for Multi-Scale Features

Yu Shi¹, Yunfeng Dong^{1,*} and Lu Tian^{2,3}

¹School of Astronautics, Beihang University, Beijing, China

²National Superior College for Engineers, Beihang University, Beijing, China

³The 15th Research Institute of China Electronic Technology Corporation, Beijing, China

*Corresponding Author: Yunfeng Dong, Email: sinosat@buaa.edu.cn

Received: 12 February 2026; Accepted: 03 May 2026; Published: 15 June 2026

ABSTRACT: Failure prognosis provides critical decision-making support for Integrated System Health Management (ISHM), ensuring the operational safety of satellites in orbit. Temporal Convolutional Networks (TCNs), known for their capability in processing time-series data, have become an important approach for failure prognosis. The gradual performance degradation of satellites, combined with multi-physics coupling effects, gives rise to multi-scale features. However, existing TCN based failure prognosis methods remain limited in their ability to simultaneously capture both local and global features, posing challenges when processing such multi-scale features. To address this issue, a Cascaded Temporal Convolution and Transformer Network (CTCTN) framework is proposed for satellite failure prognosis and uncertainty quantification. The CTCTN first adaptively aligns the feature dimensions of the Depthwise Separable Temporal Convolution (DS-TC) block and the Transformer module through Adaptive Average Pooling (AAP), enabling both local feature extraction and global dependency modeling. A heteroscedastic Huber loss function is then designed to optimize the mean and variance of the CTCTN output. Finally, epistemic and aleatoric uncertainties are separately estimated and used to construct probabilistic prediction intervals. A satellite model is developed, and a run-to-failure dataset is constructed to validate the proposed CTCTN framework using a performance degradation scenario caused by damage to the Solar Array Paddle (SAP) of a Low Earth Orbit (LEO) satellite. Experimental results demonstrate that the proposed CTCTN method not only achieves more accurate Remaining Useful Life (RUL) predictions but also effectively quantifies uncertainty arising from multi-scale features. This work provides a reference case for failure prognosis in LEO satellites and offers decision support for ISHM.

KEYWORDS: Failure prognosis; temporal convolution; transformer; uncertainty quantification; low earth orbit satellite; remaining useful life

1 Introduction

Failure prognosis is a key element in supporting decision-making for Integrated System Health Management (ISHM) [1], as it aims to prevent mission capability loss resulting from performance degradation of satellite components. As space activities continue to intensify in Low Earth Orbit (LEO), the impact of space debris has correspondingly escalated [2–4]. Among satellite subsystems, the Solar Array Paddle (SAP) could be damaged by space debris, leading to a degradation in electrical power generation [5–8]. By estimating the Remaining Useful Life (RUL) of components [9], failure prognosis enables ISHM frameworks to implement corrective measures and maintain satellite reliability.

Current research on failure prognosis primarily falls into two categories: model-driven methods and data-driven methods [9]. Model-driven approaches rely on prior knowledge to model the degradation process. Muthusamy and Kumar [10] proposed a method combining a general path model with Bayesian updating to dynamically predict the RUL of Control Moment Gyros (CMGs). Park et al. [11] employed an adaptive extended Kalman filter for failure prognosis of a reaction wheel motor. However, model-driven approaches are highly dependent on expert knowledge and are often limited by the complexity of parameter estimation.

In recent years, data-driven failure prognosis approaches have attracted growing interest due to their ability to evaluate system health status by extracting degradation trends from historical operational data [12,13]. Machine learning constitutes a primary technique in data-driven prognosis, encompassing techniques such as Random Forests (RFs) [14–16] and Support Vector Machine (SVM) [17–19]. Khelif et al. [20] applied support vector regression to model the direct relationship between sensor values and Health Indicators (HIs), allowing RUL prediction for turbofan engines throughout the degradation process. Empirical Mode Decomposition (EMD) has been adopted to preprocess bearing degradation signals, thereby facilitating SVM based RUL estimation [21]. Despite these advances, conventional machine learning techniques often struggle to process high dimensional features. In contrast, deep learning excels at extracting temporal patterns from complex degradation data [22,23], including models based on Recurrent Neural Networks (RNNs) [24] and Convolutional Neural Networks (CNNs) [25].

RNN captures temporal dependencies in time-series data by incorporating recurrent connections within their hidden layers. Long Short-Term Memory (LSTM) networks build upon standard RNN architectures by incorporating gating mechanisms that facilitate long-term information retention. Che et al. [26] developed an LSTM based recurrent network to extract patterns from sensor data and validated its effectiveness for RUL prediction using NASA's engine dataset. Wu et al. [27] applied an LSTM based model for engine performance prognosis. Sirajul Islam and Rahimi [28] developed a two-step LSTM method for reaction wheels failure prognosis. Isbilen et al. [29] proposed a hybrid algorithm that integrates LSTM networks with similarity based techniques, achieving superior performance in RUL prediction accuracy. Nevertheless, their study also highlighted the high computational demands associated with LSTM architecture.

In comparison, CNN identifies local features in historical data through sequential convolution and pooling operations, making them well suited for RUL prediction tasks. Building on this, the Temporal Convolutional Network (TCN), originally inspired by WaveNet [30], substantially enlarges the temporal receptive field, enabling models to learn long-scale temporal patterns. Chen et al. [31] transformed time-frequency features of historical data into HI and introduced the Bayesian optimization based adversarial TCN for RUL prediction. Deng et al. [32] constructed a multi-scale TCN by stacking multi-scale dilated causal convolution residual block, enabling the learning of multi-scale features. Nevertheless, TCN based architectures still struggle to process both local and global features simultaneously, making them insufficiently adapted for satellite failure prognosis.

Attention-based hybrid methods demonstrate superior capabilities in processing complex features. Hsu et al. [33] enhanced time-series feature extraction by serially combining TCN with LSTM networks and attention layers. Liu et al. [34] introduced a network with parallel TCN and LSTM branches integrated with a Convolutional Block Attention Module (CBAM) to address the performance limitations of models that consider only short-term or long-term dependencies. Zou and Lin [35] achieved multi-timescale feature extraction for aircraft engines by fusing information from multiple TCN blocks via a multi-channel attention mechanism. Furthermore, Transformer-based hybrid architectures [36,37], which benefit from the global dependency modeling capability of multi-head attention, are attracting growing research interest. Although

these methods have shown strong performance in RUL prediction for industrial equipment, there are still many challenges for satellite failure prognosis.

The main challenge for current failure prognosis research in satellites is effectively extracting multi-scale features. Owing to the stringent reliability requirements in satellite design, component performance degradation typically occurs slowly and progressively, leading to long-term degradation process. Furthermore, satellite systems are highly complex and strongly coupled [1,38]. Historical data is influenced by various factors, such as orbital dynamics and attitude maneuvers, each characterized by different temporal periods. These multi-scale features pose significant difficulties for current prognosis methods. The SAP serves as the primary power supply component by converting solar energy into electrical power. Unlike other components, the SAP is directly exposed to the harsh space environment and is therefore more vulnerable to damage from space debris impacts, which can degrade its power generation capacity [39–41]. For LEO satellites, attitude maneuvers typically occur on the order of minutes, whereas the orbital period is more than ten times longer. In contrast, the degradation process of the SAP unfolds over a much longer timescale. Consequently, in LEO satellite SAP degradation scenarios, the combined effects of orbital dynamics and attitude maneuvers on historical data generate multi-scale features, creating significant challenges for satellite failure prognosis.

To address these challenges, this paper proposes a Cascade Temporal Convolution and Transformer Network (CTCTN) for satellite failure prognosis and uncertainty quantification. Given the gradual degradation process and the presence of multi-scale features, network architectures relying solely on either TCNs or Transformers are insufficient to satisfy predictive performance requirements. To overcome this limitation, a network architecture is designed in which Transformer modules are cascaded after Depthwise Separable Temporal Convolution (DS-TC) blocks and connected through Adaptive Average Pooling (AAP). This architecture exploits the DS-TC blocks' strengths in local feature extraction and dimensionality reduction, while simultaneously enhancing the Transformer module's ability to model global dependencies. Furthermore, multi-scale features introduce uncertainty into prediction outcomes. To quantify this uncertainty, a heteroscedastic Huber loss function is designed that incorporates variance-based uncertainty constraints to jointly optimize the two output branches of the CTCTN, namely the predictive mean and predictive variance. Subsequently, Monte Carlo Dropout (MCD) is employed to separately estimate epistemic and aleatoric uncertainties based on the predicted mean and variance. To validate both the prediction accuracy and uncertainty quantification capability of the proposed CTCTN method, a satellite model is developed, and a run-to-failure dataset involving SAP degradation in a LEO satellite is constructed.

The main contributions of this paper are summarized as follows.

1. To address the challenge of multi-scale feature extraction in satellite failure prognosis, a CTCTN method is proposed. The preceding DS-TC blocks are responsible for local feature extraction and dimensionality reduction, while the subsequent Transformer modules model global dependencies. AAP enables the adaptively cascading of these two components, preserving multi-scale information and supporting progressive feature fusion from local to global representations.
2. Target optimizations are incorporated into the CTCTN architecture to enhance training efficiency and stability for long sequences, multi-parameter networks. Depthwise separable convolutions reduce the number of training parameters while maintaining effective temporal feature processing capability, thereby improving training efficiency. Furthermore, replacing the Rectified Linear Unit (ReLU) with the Gaussian Error Linear Unit (GELU) offers smooth activation properties that are particularly beneficial for stabilizing Transformer training.
3. An uncertainty quantification framework is developed for satellite failure prognosis that jointly accounts for both epistemic and aleatoric uncertainties. The CTCTN employs dual output branches to predict

the mean and variance and a heteroscedastic Huber loss function is designed to optimize these outputs. MCD is then applied to the dual-branch network outputs to separately estimate the epistemic and aleatoric uncertainties.

The remainder of this paper is organized as follows: [Section 2](#) defines the problem and the evaluation criteria for satellite failure prognosis. [Section 3](#) describes the proposed CTCTN framework for satellite failure prognosis and uncertainty quantification. [Section 4](#) presents a case study for simulation and analyzes the results. Finally, [Section 5](#) concludes the work and suggests some future directions.

2 Description of the Satellite Failure Prognosis Problem

This section formulates the satellite failure prognosis problem and describes the performance metrics used to evaluate the proposed prediction methods. Satellites operate in a highly complex space environment, where onboard components are susceptible to performance degradation caused by factors such as radiation exposure and collisions with space debris. Such degradation can ultimately impair a satellite's ability to perform mission-critical functions. As illustrated in [Fig. 1](#), a component's life cycle generally consists of two stages: a healthy stage and a degradation stage.

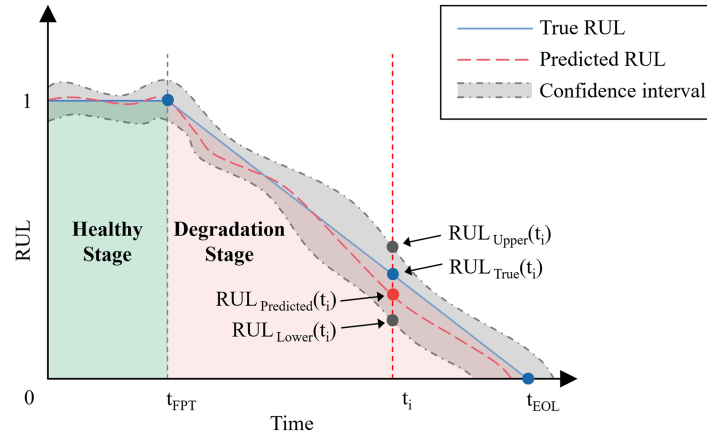


Figure 1: True RUL in the component's life cycle and the predicted RUL with confidence interval.

The satellite failure prognosis problem aims to predict the RUL of a satellite at time t_i based on historical observations, while simultaneously estimating prediction confidence intervals between $RUL_{Upper}(t_i)$ and $RUL_{Lower}(t_i)$ through uncertainty quantification. As shown in [Fig. 1](#), the solid blue line represents the true RUL, the red dashed line indicates the predicted RUL, and the gray shaded region highlights the confidence interval of the prediction. The RUL of the component is defined as follows:

$$RUL_{t_i} = \begin{cases} 1, & t_i \leq t_{FPT} \\ 1 - \frac{t_i - t_{FPT}}{t_{EOL} - t_{FPT}}, & t_i > t_{FPT} \end{cases} \quad (1)$$

where the First Prediction Time (FPT) [42] denotes the moment at which a component's performance begins to degrade, whereas the End of Life (EOL) defines the point at which the RUL reaches zero.

During the healthy stage, RUL remains fixed to one, highlighting that the component operates at full performance. Based on t_{FPT} , RUL decreases monotonically as deterioration progresses. To mitigate satellite failures, component- and mission-specific RUL thresholds are integrated to trigger predefined emergency protocols prior to critical functionality loss.

Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are used to quantitatively evaluate prediction accuracy. Both metrics measure the deviation between $RUL_{\text{Predicted}}(t_i)$ and $RUL_{\text{True}}(t_i)$, as defined below:

$$MAE = \frac{1}{n} \sum_{i=0}^n |RUL_{\text{Predicted}}(t_i) - RUL_{\text{True}}(t_i)| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^n [RUL_{\text{Predicted}}(t_i) - RUL_{\text{True}}(t_i)]^2} \quad (3)$$

where lower RMSE and MAE values indicate better predictive accuracy.

In addition, two other metrics are employed to evaluate the uncertainty intervals: Prediction Interval Coverage Probability (PICP), which evaluates the accuracy of the predicted interval, and Mean Prediction Interval Width (MPIW), which quantifies the uncertainty. These metrics are expressed as follows:

$$PICP = \frac{1}{n} \sum_{i=0}^n f(t_i) \quad (4)$$

$$MPIW = \frac{1}{n} \sum_{i=0}^n [RUL_{\text{Upper}}(t_i) - RUL_{\text{Lower}}(t_i)] \quad (5)$$

where $f(t_i)$ represents a flag indicating if the true value lies within the prediction interval. If the true value falls outside $RUL_{\text{Upper}}(t_i)$ or $RUL_{\text{Lower}}(t_i)$ boundaries, $f(t_i) = 0$; otherwise, $f(t_i) = 1$. A narrower interval, reflected by a lower MPIW, typically results in a reduced PICP. Therefore, a combined analysis of both metrics provides an assessment of the method's uncertainty quantification performance.

When combined together, these four metrics (e.g., RMSE, MAE, PICP, and MPIW) form a comprehensive framework for assessing the performance of the prognosis method.

3 Methods

To address the challenges of failure prognosis and uncertainty quantification associated with SAP degradation in LEO satellites, this paper proposes the CTCTN framework. Stacked DS-TC blocks are first cascaded with Transformer modules through AAP, enabling effective local feature extraction and global dependency modeling for multi-scale features. The heteroscedastic Huber loss function is then designed during training to optimize both the mean and variance outputs of CTCTN. Moreover, MCD is applied during testing to quantify epistemic and aleatoric uncertainties. Finally, the framework generates predictions for the RUL along with their corresponding prediction intervals.

3.1 The Structure of Cascaded Temporal Convolution and Transformer Network (CTCTN)

The proposed CTCTN structure consists of four key parts: DS-TC block, AAP, Transformer module, Multi-Head Attention Pooling (MHAP). These parts collectively facilitate hierarchical feature extraction and cross-scale fusion, enhancing the representation of degradation patterns.

The DS-TC block preserves the temporal modeling capabilities of conventional TCN while improving computational efficiency through the depthwise separable convolutions. Serving as the foundational unit of the DS-TC block, the dilated causal convolution (DCC) expands the receptive field without increasing the kernel size by adjusting the dilation rate. The receptive field comparison with standard convolution is shown in Fig. 2.

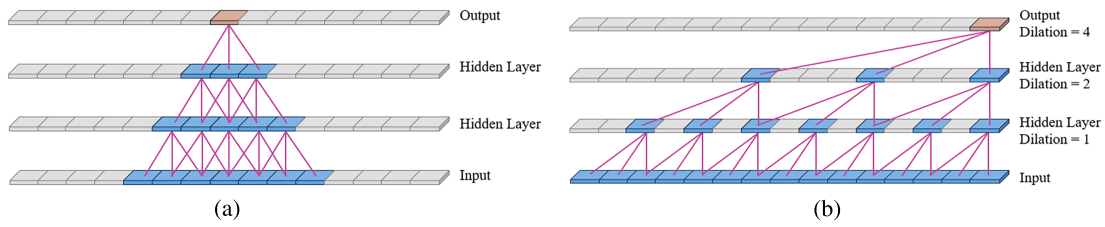


Figure 2: The receptive field: (a) standard convolution; (b) dilated causal convolution.

The causal convolution structure enables the network to learn historical information in the temporal data and avoid the effect of future information. Notably, this study incorporates depthwise and pointwise convolutions into the DCC, resulting in a depthwise separable convolution structure, as illustrated in Fig. 3.

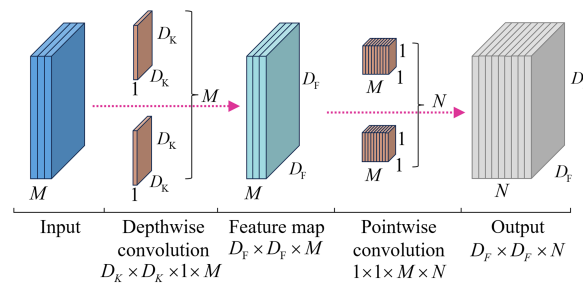


Figure 3: The structure of depthwise separable convolution.

Under identical input-output conditions, the computational cost of traditional convolution is given by $D_c = D_F^2 \times D_K^2 \times M \times N$, whereas the cost of depthwise separable convolution is expressed as $D_{DSC} = D_{Depth} + D_{Point} = D_F^2 \times D_K^2 \times M + D_F^2 \times M \times N$. The computational cost ratio between the two convolution types is $1/N + 1/D_K \times D_K$, demonstrating that depthwise separable convolution effectively reduces the computational burden. Following this, batch normalization and activation functions are incorporated via residual skip connections to form a DS-TC block, as illustrated in Fig. 4. A depthwise separable convolution consists of a depthwise convolution followed by a pointwise convolution. In this study, the depthwise convolution is implemented as a DCC. By stacking i DS-TC blocks with dilation rates set to 2^i in each block, multi-scale feature extraction is achieved.

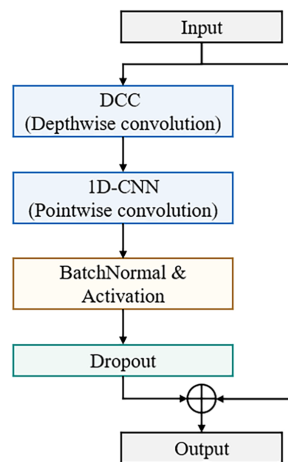


Figure 4: The structure of the proposed DS-TC block.

A dimension mismatch exists between the output features of the DS-TC blocks and the input features of the cascaded Transformer modules, which is addressed by AAP. Specifically, the length of the input temporal sequences varies depending on the task, leading to a corresponding change in the output feature dimension of the DS-TC blocks. Additionally, the number of stacked layers in both the DS-TC blocks and Transformer modules, as well as the parameters within the network, contribute to the mismatch in feature dimensions. Considering that the output feature dimension of the DS-TC block is (B, C, T) , where B denotes batch size, C represents the number of channels, and T stands the time steps, the AAP is used. It performs global average pooling across the T time steps for each channel, downsampling the temporal dimension to a fixed length $D_{\text{Transformer}}$ which matches the input dimension of the Transformer module. Therefore, the cascading mechanism in this paper is expressed as $H_{\text{Transformer}} = \text{AAP}(H_{\text{DS-TC}})$. The output of the DS TC block $H_{\text{DS-TC}}$ is connected to the input of the Transformer module $H_{\text{Transformer}}$ through AAP. Moreover, global average pooling not only preserves the key features of the channels but also suppresses noise interference, which is particularly advantageous for processing data from sensors exposed to space environments.

The Transformer module receives features from the AAP to capture global dependencies. The structure of Transformer module is shown in Fig. 5. The DS-TC block reduces the dimension of the temporal sequence, thereby reducing the computational cost of the Transformer module. Furthermore, AAP further aligns the feature dimensions, ensuring compatibility with the input requirements of the Transformer module. Initially, the features are added to the Transformer module through positional encoding. It encodes the positional information of the input sequence, enabling the network to learn how relative positions affect the structure, thus allowing it to extract temporal features. Global dependencies across different timesteps within the features are then captured through stacked Transformer encoders. The self-attention mechanism in Transformer enables each time step to attend to all other steps simultaneously, breaking the limitation of local receptive fields in traditional CNN or RNN. This is particularly beneficial for satellite data, where long-period events (e.g., orbital period) require modeling dependencies across many time steps. In multi-head attention setup, each head can learn to focus on different temporal scale, enabling the Transformer to process multi-scale features present in satellite failure scenarios. Notably, the activation function used in the Transformer encoders in this study is GELU which is preferred in attention-based Transformer architectures due to its beneficial properties, including nonlinearity, differentiability, and smoothness. These unique properties contribute significantly to training stability of the network. The output features of the stacked Transformer encoders in Fig. 5 form a tensor of shape $(B, D_{\text{Transformer}}, C_{\text{Transformer}})$, where B denotes the batch size, $D_{\text{Transformer}}$ the sequence length, and $C_{\text{Transformer}}$ the feature dimension determined by the Transformer's hidden dimension. This representation preserves global contextual information and is subsequently fed into the MHAP module for further feature aggregation.

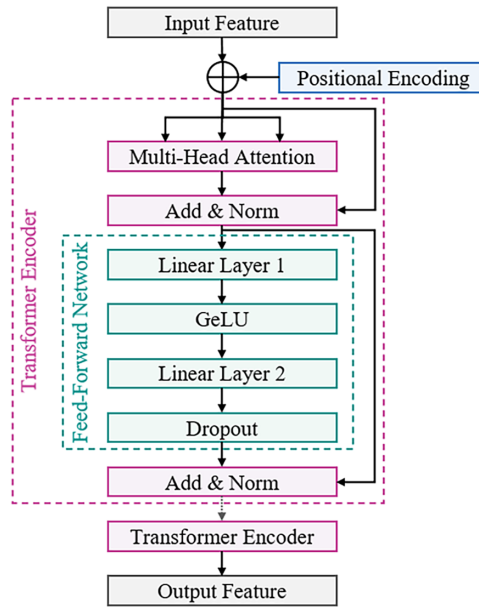


Figure 5: The structure of the Transformer module.

MHAP extracts multi-dimensional features from the Transformer output and a following fully connected layer outputs predicted mean and variance of RUL. Unlike conventional average or max pooling, it introduces a learnable query vector to compute a weighted sum over the entire sequence, leveraging the multi-head attention mechanism for enhanced feature fusion. Specifically, the input sequence \mathbf{x} and the expanded learnable query vector are linearly projected as follows.

$$\begin{cases} \mathbf{Q}_h = \mathbf{query} \cdot \mathbf{W}_h^Q \\ \mathbf{K}_h = \mathbf{x} \cdot \mathbf{W}_h^K \\ \mathbf{V}_h = \mathbf{x} \cdot \mathbf{W}_h^V \end{cases} \quad (6)$$

where \mathbf{W}_h^Q , \mathbf{W}_h^K , and \mathbf{W}_h^V are trainable weight matrices that project the input into query, key, and value representations for each individual attention head. The attention scores for each head are normalized as follows.

$$\mathbf{Attention}_h = \text{softmax} \left(\mathbf{Q}_h \mathbf{K}_h^T / \sqrt{d_K} \right) \quad (7)$$

where d_K donates the dimension of the key vectors, which is used to prevent gradient explosion. The score matrix encodes the similarity weights between the query and each position in the sequence. Notably, the attention distribution learned by each head differs, incorporating multi-dimensional degradation features. Subsequently, the value vectors are weighted and combined using the attention scores to generate the output vector for each head, as follows.

$$\mathbf{Output}_h = \mathbf{Attention}_h \cdot \mathbf{V}_h \quad (8)$$

finally, the concatenated outputs from all heads are integrated into the attention output through a linear layer, as follows.

$$\mathbf{Output} = \text{Concat} (\mathbf{Output}_1, \mathbf{Output}_2, \dots, \mathbf{Output}_H) \cdot \mathbf{W}^O \quad (9)$$

where W^O is the output weight matrix. The output features are subsequently passed through a fully connected layer, yielding a two-dimensional vector representing the predicted mean and variance of RUL.

3.2 Epistemic and Aleatoric Uncertainty Quantification

To quantify the uncertainty arising from multi-scale features, this paper proposes a method framework for quantifying both epistemic and aleatoric uncertainty, comprising two stages: network training and testing. During training, the outputs of the CTCTN are used as inputs to a heteroscedastic Huber loss function for network parameter optimization. During testing, epistemic and aleatoric uncertainties are then quantified through the MCD method to calculate confidence intervals.

In the training stage, the heteroscedastic Huber loss function is designed to optimize the the mean μ and variance σ^2 output branches of the CTCTN. The network's prediction is assumed to follow a Gaussian distribution: $\mathbf{p}(\mathbf{y}|\mathbf{x}) = N[\mu(\mathbf{x}), \sigma^2(\mathbf{x})]$, where \mathbf{x} denotes the input data and \mathbf{y} corresponds to the predicted output [43,44]. To train this dual-output network, a heteroscedastic Huber loss function is given as follows.

$$\begin{aligned} \mathbf{L}_{\text{total}} &= \mathbf{L}_{\text{huber}} + \alpha \mathbf{L}_{\text{uncertainty}} \\ \mathbf{L}_{\text{huber}} &= 1/N \sum_{i=1}^N \begin{cases} 1/2 (\mu_i - y_i)^2, & |\mu_i - y_i| < \delta \\ \delta (|\mu_i - y_i| - 1/2\delta^2), & |\mu_i - y_i| \geq \delta \end{cases} \\ \mathbf{L}_{\text{uncertainty}} &= 1/N \sum_{i=1}^N \left\{ 1/2 \left[\sigma_i^2 + (\mu_i - y_i)^2 / \sigma_i^2 \right] + 1/10 \ln(1 + \sigma_i^2) \right\} \end{aligned} \quad (10)$$

where $\mathbf{L}_{\text{huber}}$ is the Huber loss and $\mathbf{L}_{\text{uncertainty}}$ is a variance-based uncertainty constraint. The Huber loss [45], tailored for regression tasks, integrates the characteristics of both mean squared error and mean absolute error, thereby mitigating noise sensitivity and improving model robustness. The uncertainty constraint includes a variance calibration term $1/2 \left[\sigma_i^2 + (\mu_i - y_i)^2 / \sigma_i^2 \right]$ and a variance regularization term $1/10 \ln(1 + \sigma_i^2)$. The variance calibration term consists of the variance and the standardized residual $(\mu_i - y_i)^2 / \sigma_i^2$. The predicted variance directly limits variance overestimation, while the residual term associates prediction errors with variance. The variance regularization term contains the natural logarithmic function to restrict excessive growth of variance and avoids instability when variance approaches zero. Network parameters are optimized using the AdamW algorithm.

In the testing stage, MCD is used to quantify predictive uncertainty. Originally introduced as a regularization method to avoid overfitting, dropout randomly deactivates neurons during training, thereby reducing reliance on specific network pathways. MCD extends this strategy by maintaining dropout activation in the testing stage. By executing multiple stochastic forward passes, it produces a distribution over predictions, which facilitates the estimation of uncertainty.

Predictive uncertainty is quantified using the mean and variance of the outputs from MCD. This uncertainty consists of two components: epistemic and aleatoric uncertainty [46,47]. Epistemic uncertainty stems from the variability of network parameters (e.g., weights) and reflects the uncertainty of the network itself. It arises when limited training data prevent accurate estimation of optimal parameters. Aleatoric uncertainty is attributed to noise from the data acquisition process (e.g., sensor error, environmental disturbance) and is not related to the model parameters. Even when the network parameters are fixed, aleatoric uncertainty persists due to randomness in the outputs. The total predictive uncertainty is given by:

$$\text{Var}(\mathbf{y}) = \text{Var}[\mathbf{E}(\mathbf{y}|\mathbf{x})] + \mathbf{E}[\text{Var}(\mathbf{y}|\mathbf{x})] \quad (11)$$

where $\text{Var}[E(\mathbf{y}|\mathbf{x})]$ and $E[\text{Var}(\mathbf{y}|\mathbf{x})]$ represent epistemic and aleatoric uncertainty, respectively [48]. Epistemic uncertainty arises from variations in network parameters, which induce fluctuations in the expectation $E(\mathbf{y}|\mathbf{x})$. This variability is quantified by $\text{Var}[E(\mathbf{y}|\mathbf{x})]$, reflecting the intrinsic uncertainty within the network. In contrast, aleatoric uncertainty, expressed as $E[\text{Var}(\mathbf{y}|\mathbf{x})]$, directly captures the variance due to data noise. The 95% Confidence Interval (CI) [49] for predictions can be expressed as.

$$CI = E(\mathbf{y}|\mathbf{x}) \pm 1.96 \cdot \text{Var}(\mathbf{y}) \quad (12)$$

3.3 The CTCTN Framework for Satellite Failure Prognosis and Uncertainty Quantification

To address the challenge of multi-scale feature extraction in SAP degradation for LEO satellites, this paper proposes a CTCTN framework for satellite failure prognosis and uncertainty quantification, as illustrated in Fig. 6. The proposed framework consists of three main components: scenario configuration and data processing, prediction network construction, and uncertainty quantification.

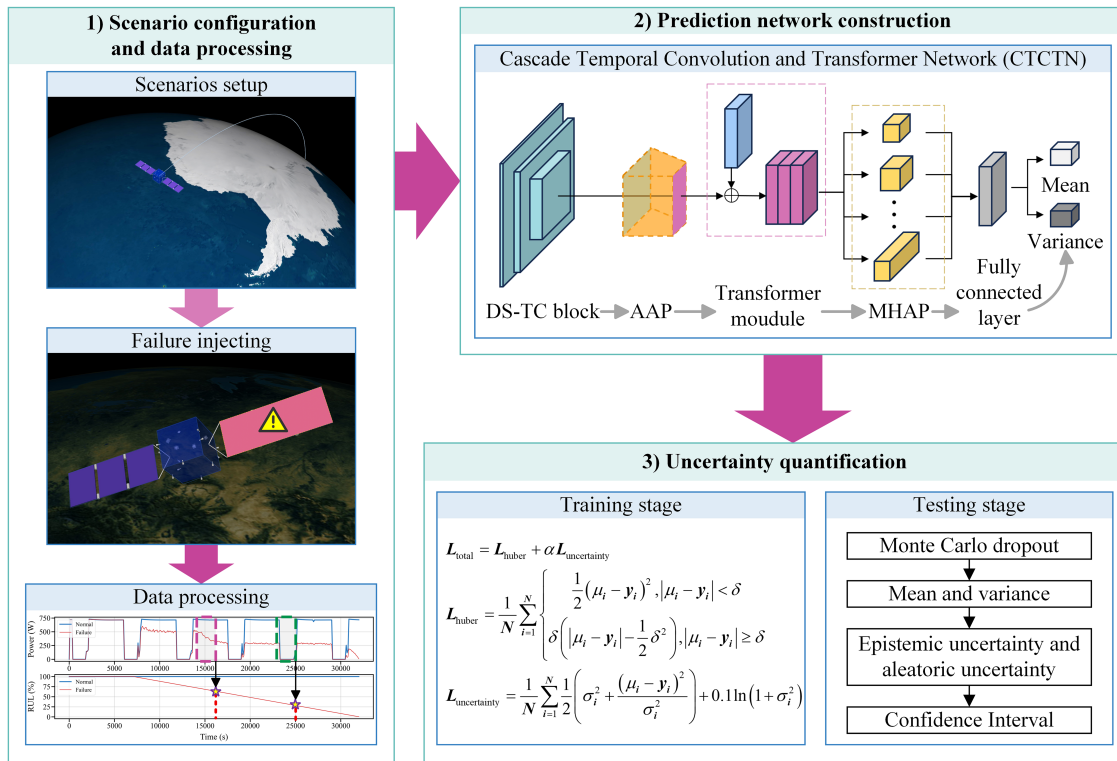


Figure 6: The CTCTN framework for satellite failure prognosis and uncertainty quantification.

In the first component, a LEO satellite simulation scenario is developed using a digital satellite development platform [50,51]. Subsequently, failures are injected into the satellite model during simulation to generate run-to-failure data, from which a satellite run-to-failure dataset is constructed under diverse operational conditions.

In the second component, the CTCTN architecture is proposed by combining temporal convolution with attention mechanisms to extract multi-scale degradation features from historical data. The network stacks DS-TC blocks to achieve local feature extraction and dimensionality reduction. An AAP then adaptively aligns feature dimensions, facilitating cascading of DS-TC blocks and Transformer modules. The Transformer module utilizes self-attention mechanisms to perform global contextual modeling of temporal

features, establishing dependencies across different temporal scales and highlighting critical degradation patterns. By incorporating learnable query vectors with a multi-head attention, MHAP facilitates cross-scale feature interactions. Finally, fully connected layers are used to output the predictive mean and variance of the RUL.

In the third component, both epistemic and aleatoric uncertainties are quantified. During training, a heteroscedastic Huber loss function is employed to optimize the network parameters by jointly learning the predictive mean and variance of the RUL. The Huber loss function computes residuals using a piecewise structure. The uncertainty constraint processes predictive variance through a variance calibration term and a variance regularization term. During testing, the MCD method is applied to quantify both epistemic and aleatoric uncertainties. The proposed framework ultimately produces predictive intervals for RUL.

4 Experiments and Results

The failure prognosis performance of the proposed CTCTN method is evaluated using an SAP degradation scenario for a LEO satellite. Firstly, a satellite model is developed based on a digital satellite development platform [50,51] which enables the construction of C++ satellite model code frameworks through standardized configuration files [38]. This satellite model is used to simulate the power generation degradation process resulting from SAP damage caused by space debris impacts [5–7,41,52]. The multi-scale features of the degradation process are analyzed by comparing the timescales of orbital dynamics, attitude maneuvers, and degradation process. Secondly, a dataset comprising run-to-failure data under various operational conditions is constructed to evaluate the predictive performance of the proposed CTCTN method for the SAP degradation scenario. An optimal set of hyperparameters is selected via grid search based on both predictive performance and uncertainty quantification performance. Ablation experiments are then conducted to validate the effectiveness of the CTCTN design. Finally, the proposed CTCTN is compared against current failure prognosis methods, demonstrating its superior predictive performance for the satellite SAP degradation scenario.

The failure scenario simulation and prognosis are performed using a Dell workstation equipped with Intel[®] Core™ i7-8700 CPU @ 3.20 GHz, 16 GB of RAM, and an NVIDIA GeForce GTX 1050 Ti GPU with 4 GB of VRAM. The software configuration of the simulation and prognosis are C++ and Python, respectively.

4.1 Dataset Description

A LEO satellite model is developed to generate run-to-failure datasets under various operational conditions. Space debris poses a serious threat to satellite safety, and a damage to SAP can lead to significant degradation of power generation capacity [39,40]. Accurate RUL prediction for satellites SAP degradation is therefore critical [53]. Accordingly, this study develops a satellite model to simulate the power output degradation process of a damaged SAP. The satellite model integrates multiple subsystems, including attitude and orbit control, power, thermal control, propulsion, payload, and structure subsystem. For more model details, readers can refer to references [54,55]. The attitude and orbit control subsystem governs orbital and attitude maneuvers, while the power and thermal control subsystems manage energy and thermal balance, respectively. In the healthy state, the output power of the SAP is expressed as follows:

$$P_{wing} = C_{shadow} S_0 A_{wing} \eta X_{wing} \cos(\theta_{SW}) (\beta_P \Delta T_{wing} + 1) \quad (13)$$

where C_{shadow} identifies whether the solar wings are in the Earth's shadow, S_0 quantifies the solar irradiance, A_{wing} represents the area of the photovoltaic array on the solar wings, η denotes the photoelectric conversion efficiency of the array, θ_{SW} describes the angle between the normal vector of the solar wings

and the sunlight's direction, β_P is the power temperature coefficient of the solar wings, X_{wing} encompasses additional coefficients, and ΔT_{wing} indicates the temperature differential from the standard temperature of the solar wings.

At a random time during the simulation, a failure is injected, after which the output power of the SAP begins to degrade. The injected parameter is the degradation duration $t_{duration}$. Both the failure injection time and the degradation duration are randomized to generate diverse operating conditions for network training and to enhance its generalization capability. Following fault injection, the degraded output power of the SAP is expressed as $P_{wing,degradation} = P_{wing} \cdot e^{-\lambda \Delta t_{degradation}}$, where $\lambda = k/t_{duration}$, $\Delta t_{degradation}$ represents the simulation time of the degradation process, and k denotes the degradation coefficient. To account the uncertainty in the degradation process, a multiplicative uniform noise of $\pm 5\%$ is introduced to the degradation coefficient.

The initial parameters for the satellite model based on the real LEO satellite (referred to as SL) are listed in Table 1.

Table 1: The initial configuration parameters for the satellite model.

Parameter	Value
Orbit elements	{7,018,455 m, 0.002, 97.948°, 284.766°, 217.387°, 142.619°}
Attitude	{0°, 0°, 0°}
Charge voltage (V)	28.324
Capacity (mAh)	70,000
SAP area (m ²)	6
SAP photoelectricity conversion efficiency	0.3
SAP power-temperature coefficient	-0.0002
SAP Max Output Power (W)	1260
Initial time (UTC)	2019-9-22 00:00:00

Fig. 7 presents a comparison between the telemetry data of SL and the simulation data generated by the satellite model. In Fig. 7, the solid line represents the telemetry data, while the marked line denotes the simulation data. It can be observed that the orbital and attitude of the satellite model are consistent with telemetry data. Moreover, the trends of component parameters—such as current, voltage, and temperature—also exhibit strong agreement. These results confirm that the operating condition of the satellite model are consistent with SL.

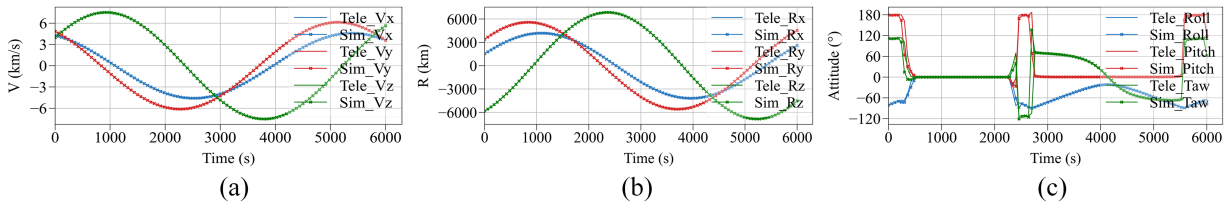


Figure 7: (Continued)

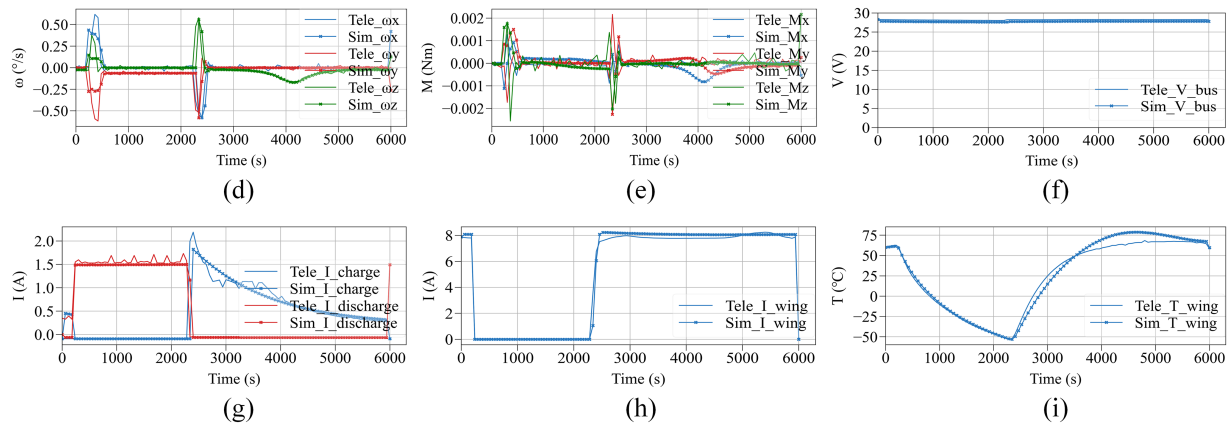


Figure 7: Comparison between telemetry data and simulation data: (a) position; (b) velocity; (c) attitude angular; (d) attitude angular velocity; (e) torque of the wheel; (f) bus voltage; (g) battery charging and discharging current; (h) SAP current; (i) SAP temperature.

The multi-scale features in the degradation process are analyzed using the simulation data shown in Fig. 8, under both normal and degraded states. In Fig. 8, the blue line represents the normal operational data, while the red line corresponds to the run-to-failure data. Moreover, the t_{EOL} at 32,100 s and the t_{FPT} at 7000 s are both marked using red dashed lines. The orbital period, highlighted by a blue box, is 6000 s. A 2000-s window, highlighted by the green box, of SAP output power before the moment t_i (marked by a yellow star) is used as a sample, with the RUL at t_i serving as the label.

As shown in Fig. 8, the output power of SAP exhibits progressive degradation following fault injection. This simulation is based on the performance degradation process of a damaged SAP described in previous studies [5,7,52]. During the initial degradation stage, the limited extent of damage has only a minor impact on SAP power generation. The power subsystem is still able to maintain energy balance, and battery capacity remains largely unchanged. Subsequently, as the severity of degradation increases, the energy balance is disrupted, leading to an accelerated battery consumption. Due to the continuously changing relative positions of the satellite, Earth, and Sun, the satellite periodically enters Earth's shadow. During these eclipse periods, the SAP cannot perform photoelectric conversion, causing its power output to drop to zero. The satellite also adjusts its attitude to an Earth-pointing mode during this time. It is noteworthy that the RUL cannot be inferred from the power data during eclipse, introducing uncertainty into the prediction. After exiting the eclipse, the satellite readjusts its attitude to a Sun-pointing mode, and the SAP power gradually recovers. Compared to the orbital period, the satellite's attitude maneuver occurs rapidly, while SAP degradation is a slow process. These factors give rise to multi-scale features in the SAP power degradation process.

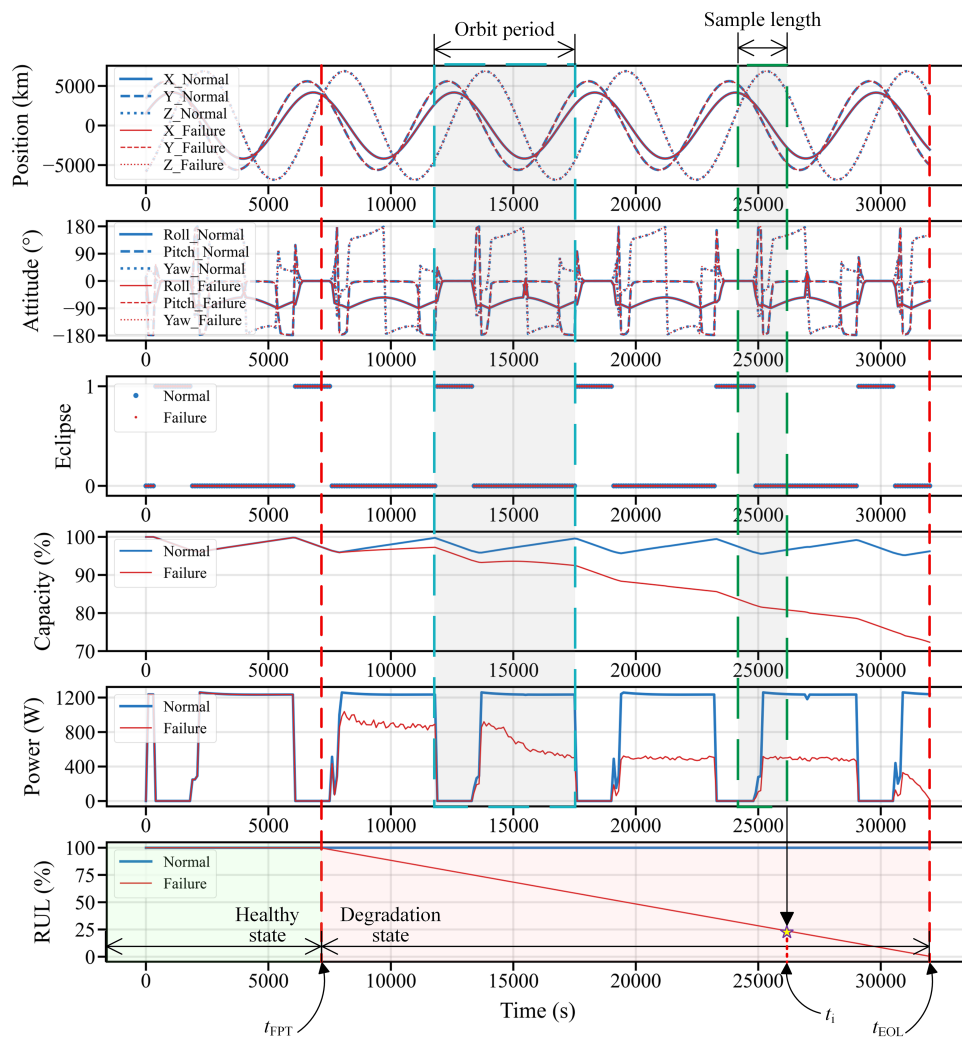


Figure 8: The simulation data in normal and SAP degradation scenario.

The satellite run-to-failure dataset comprises simulation data from 10 operational conditions, with the specific failure scenario parameters detailed in Table 2. The simulation model records data at a frequency of 1 Hz (once per second). Starting from the 2500-s into the simulation, the SAP output power over the preceding 2000 s is collected every 5 min as a sample. A total of 1022 samples are collected, with the number of samples per operating condition detailed in Table 2. The RUL at the time of sample collection, normalized between 0 and 1 according to Eq. (1), is assigned as the label. This data processing and labeling strategy is the same as logic of real satellite failure prognosis, where the RUL at a moment is inferred from recent historical data. Each operational condition is simulated once to generate the corresponding samples. To evaluate generalization capability, a leave-one-out strategy is adopted. Specifically, one operating condition is used as the test set, while the remaining conditions constitute the training set, and the average value of the evaluation metrics across all conditions are calculated. To mitigate overfitting, 20% of the training samples are reserved as a validation set.

Table 2: The configuration of 10 operational conditions in the satellite run-to-failure dataset.

Condition Number	Failure Initiation Time (min)	Failure Duration (min)	Total Simulation Duration (min)	Sample Number
1	108	128	236	48
2	175	651	826	166
3	108	462	570	121
4	275	166	441	89
5	308	171	479	96
6	75	622	697	140
7	108	309	417	84
8	75	367	442	89
9	341	135	476	96
10	41	455	496	93

4.2 Hyperparameter

To avoid the impact of hyperparameters on prediction performance, random search is adopted for hyperparameter optimization. The search ranges and corresponding results are listed in Table 3. To reduce the risk of overfitting, a dropout rate is set to 0.2. GELU is used as the activation function to provide a smooth, non-linear activation. The AdamW optimizer is employed to minimize the loss function and update network parameters. The learning rate is set to 0.001, the batch size is set to 128, and the number of epochs is 500. α and δ are set to 0.1 and 1, respectively, in proposed heteroscedastic Huber loss function. Finally, the number of Monte Carlo iterations is set to 100.

Table 3: The optimal hyperparameters of the proposed CTCTN.

Parameter	Range	Configuration
Convolution channels	[16, 32, 64]	16
Kernel size	[3, 5, 7]	7
DS-TC layers	[1, 2, 3, 4]	2
Transformer dimensions	[16, 32, 64]	64
Heads	[4, 8, 16]	16
Transformer layers	[1, 2, 3, 4]	3

4.3 Predicted Results

The predictive performance of the proposed CTCTN is evaluated using the satellite run-to-failure dataset. Fig. 9 illustrates the predicted RUL for operational conditions 9 and 10 under the leave-one-out testing. In the upper subplot, the blue line denotes the true RUL, the red line represents the predicted RUL, and the gray-shaded area indicates the 95% confidence interval. In the lower subplot, the blue points correspond to epistemic uncertainty, the red points to aleatoric uncertainty, and the black points to total uncertainty.

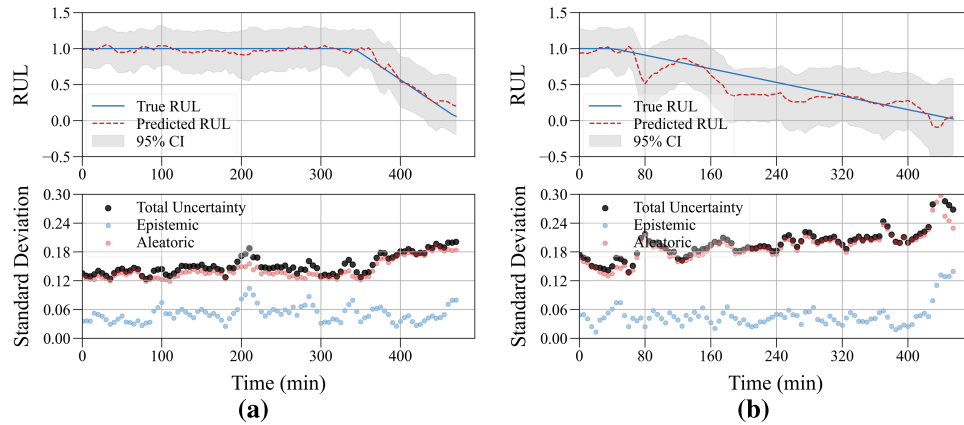


Figure 9: The predicted RUL and uncertainty quantification results: (a) operational condition 9; (b) operational condition 10.

As shown in Fig. 9, the proposed CTCTN effectively predicts the RUL of the satellite SAP. During the healthy stage, the predicted RUL closely matches the true RUL, and the associated uncertainty remains stable. In the degradation stage, both the predicted RUL and uncertainty exhibit periodic fluctuations, which are particularly pronounced under operational condition 10. These fluctuations correspond to eclipse periods, during which the SAP generates zero power, thereby affecting the RUL prediction. Despite this, the overall degradation trends of the true and predicted RUL remain consistent, and the true RUL is consistently encompassed within the predicted uncertainty interval. Furthermore, in the uncertainty quantification results, the aleatoric uncertainty is larger than the epistemic uncertainty, representing the main part of the total uncertainty. This indicates that the network has effectively learned the degradation patterns, while the multi-scale features of the satellite system primarily contribute to the aleatoric uncertainty in the predicted RUL. In summary, the proposed CTCTN framework effectively predicts the RUL of the satellite SAP and provides well uncertainty quantification.

4.4 Ablation Experiment

To validate the effectiveness of the cascaded design in CTCTN, ablation experiments are conducted on modules and connection types. Experimental configurations are listed in Table 4, where Models A and B include only a single module, while Models C, D, and E employ different connection types.

Table 4: The configuration of failure scenarios.

Model	DS-TC	Transformer	Connection Type	Description
Model A	✓	×	–	Only DS-TC
Model B	×	✓	–	Only Transformer
Model C	✓	✓	0	Parallel TC-Transformer
Model D	✓	✓	1	Cascade Transformer-TC
Model E	✓	✓	2	Cascade TC-Transformer

The prediction results of each model are presented in [Table 5](#), demonstrating the effectiveness of the proposed CTCTN structure. Firstly, compared to Model A, Model E achieves RMSE reductions of 50%, with corresponding MAE reductions of 49%. Similarly, compared to Model B, Model E reduces RMSE by 45%, and MAE by 42%. These results indicate that Models A and B, which incorporate only a single module, exhibit limited predictive capability for multi-scale degradation features, as they are restricted to processing either local or global features alone. Secondly, Model E reduces RMSE by 54% compared to Model C, with MAE reductions of 58%. Compared to Model D, Model E achieves RMSE reductions of 55%, and MAE reductions of 57%. These results demonstrate the impact of different connection types on predictive performance for the satellite SAP degradation process.

Table 5: Prediction results of the ablation experiment.

Method	RMSE		MAE		PICP		MPIW	
Model A	0.2471	(-)	0.1835	(-)	0.8813	(-)	0.7879	(-)
Model B	0.2233	(+9.63%)	0.1595	(+13.11%)	0.8494	(-3.62%)	0.7103	(+9.84%)
Model C	0.2695	(-9.06%)	0.2206	(-20.20%)	0.8318	(-5.62%)	0.7265	(+7.79%)
Model D	0.2747	(-11.16%)	0.2160	(-17.70%)	0.8213	(-6.81%)	0.6959	(+11.67%)
Model E	0.1236	(+49.98%)	0.0930	(+49.32%)	0.9829	(+11.53%)	0.6879	(+12.69%)

The impact of connection types on predictive performance is further examined. Model C shows no improvements over Model A and B. This suggests that, the parallel structure fails to fully exploit the advantages of the single module. In model D, the Transformer first transforms multi-scale features into a global representation, but the subsequent DS-TC is unable to sufficiently capture local patterns, resulting in a poor predictive performance. In contrast, the proposed cascaded design—where the DS-TC preceding the Transformer—achieves effective multi-scale feature fusion and superior prognostics. This arrangement allows the DS-TC to extract local features while performing dimensionality reduction, and simultaneously enhances the Transformer’s capacity to model long-range dependencies.

Furthermore, the capability for uncertainty quantification is evaluated using the PICP and MPIW metrics. Model E achieves the highest PICP values and the narrowest MPIW. This indicates that the proposed CTCTN structure not only provides the most reliable coverage of uncertainty intervals but also produces the tightest interval widths, demonstrating superior uncertainty quantification performance for satellite failure prognosis.

A comparative analysis is further conducted to evaluate performance differences among different modules integrated after the Cascade TC-Transformer. The results are summarized in [Table 6](#). Compared with the Global Average Pooling (GAP), MHAP achieves a 63% reduction in RMSE and a 69% reduction in MAE. The inferior predictive performance of GAP can be attributed to its averaging across all time steps, which limits its ability to capture multiscale feature variations. In addition, compared with the Classification (CLS) token, MHAP yields a 46% reduction in both RMSE and MAE. By leveraging the multi-head attention mechanism, MHAP exhibits a stronger capability for extracting multi-scale features than the CLS token. Consequently, MHAP is better suited for the satellite fault prognosis task.

Table 6: Prediction results of CTCTN with different module.

Method	RMSE	MAE	PICP	MPIW
Cascade TC-Transformer with CLS token	0.3336	0.2987	0.9210	1.0269
Cascade TC-Transformer + GAP	0.2287	0.1708	0.8598	0.7142
Cascade TC-Transformer + MHAP	0.1236	0.0930	0.9829	0.6879

4.5 Comparison Experiment

To further validate the performance of the proposed CTCTN for satellite failure prognosis, it is compared with current prognostic methods. The specific configurations of these comparative methods are detailed as follows, with hyperparameters kept consistent with those of the proposed method:

1. MCA-TCN [35]: The network integrates three TCN blocks with multi-channel attention for feature fusion. Each TCN block contains three dilated convolutional layers with 16 channels and a kernel size of seven. The fused features are processed by a GRU layer with 18 hidden units, followed by a fully connected layer for the RUL.
2. TCLSTM [33]: This network uses a TCN module with one layer and 16 channels, followed by an LSTM layer with 16 hidden units and an attention mechanism. The TCN kernel size is seven. The attended features are passed to a fully connected layer for output.
3. Attention-BiLSTM [56]: The network architecture consists of a two-layer BiLSTM with an attention mechanism, where each LSTM layer comprises 16 hidden units.
4. Transformer-LSTM [36]: This architecture employs a three-layer Transformer followed by a two-layer LSTM. The Transformer module is configured with 16 heads, while each LSTM layers contains 16 hidden units.
5. Informer [57]: The network employs a ProbSparse self-attention mechanism to reduce computational complexity, stacking three encoder layers with distilling operations to progressively shorten the sequence length. A global average pooling layer followed by a fully connected layer outputs the predicted mean and variance of the RUL.
6. PatchTST [58]: The model segments the input univariate time series into patches of length 16 with a stride of 8, linearly projecting each patch into a 64-dimensional representation. After processing by three Transformer encoder layers, the feature of the last patch is fed into a fully connected layer to produce the RUL mean and variance.
7. CNN-Transformer [37]: The network composed of a two-layer CNN followed by a three-layer Transformer. CNN uses 16 channels with a kernel size of seven, whereas the Transformer is configured with 16 heads.

Fig. 10 presents the prediction results of all methods under operating conditions 9 and 10, where the lines represent the true RUL and the predicted RUL of each method and the scatter plots depict the overall uncertainty. It is evident that the proposed CTCTN achieves superior predictive performance. Specifically, during the healthy state, the RUL predicted by CTCTN remains consistently close to the true RUL. In contrast, methods including CNN-Transformer and MCA-TCN yield lower RUL predictions under condition 9. Furthermore, during the degradation state, the predicted RUL from CTCTN more accurately tracks the true RUL, whereas TCLSTM, Attention-BiLSTM, Transformer-LSTM, and CNN-Transformer exhibit larger fluctuations.

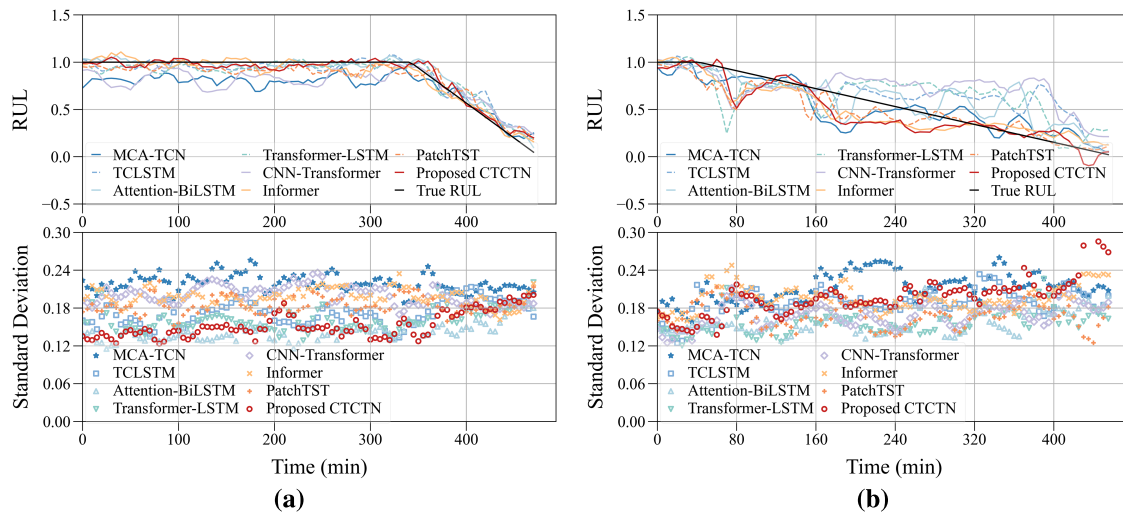


Figure 10: The predicted RUL and uncertainty quantification results of all methods: (a) operational condition 9; (b) operational condition 10.

The statistic prediction results of all methods are summarized in Fig. 11 and Table 7. To avoid the impact of randomness, 10 repetitions of the experiment are conducted and the mean and variance are reported. The findings demonstrate that the proposed CTCTN method achieves the best overall performance in terms of both prediction accuracy and uncertainty quantification. TCLSTM exploits the strengths of TCN in extracting temporal features from historical data for RUL prediction. However, their feature extraction capabilities are limited in the satellite SAP degradation scenario, which involves multi-scale features. Although TCLSTM incorporates an attention mechanism, its single-head attention cannot simultaneously capture degradation patterns across multiple time scales. A similar limitation is also observed in the attention-based MCA-TCN. The predictive performance of Transformer-LSTM is constrained by its preceding Transformer module, which is consistent with the findings of the ablation study. Although the architecture of CNN-Transformer is similar to that of the proposed CTCTN, CNNs have limited capacity in processing temporal data. Attention-BiLSTM, owing to its residual connections, effectively mitigates the vanishing gradient problem and consequently achieves relatively low RMSE. Informer and PatchTST are recent strong baselines. Compared with Informer, the proposed CTCTN reduces RMSE and MAE by 10.60% and 12.65%, respectively. Furthermore, while achieving a PICP of 98.48%, the proposed CTCTN also yields a relatively low MPIW. In summary, the proposed CTCTN effectively captures multi-scale features in satellite SAP degradation data, leading to state-of-the-art predictive performance and superior uncertainty quantification.

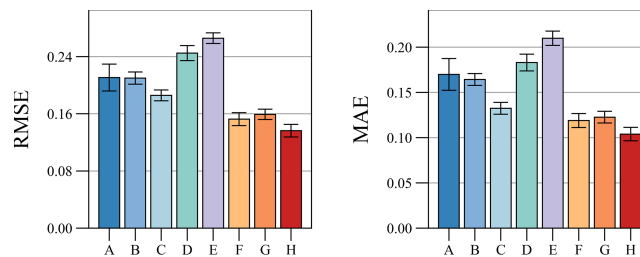


Figure 11: (Continued)

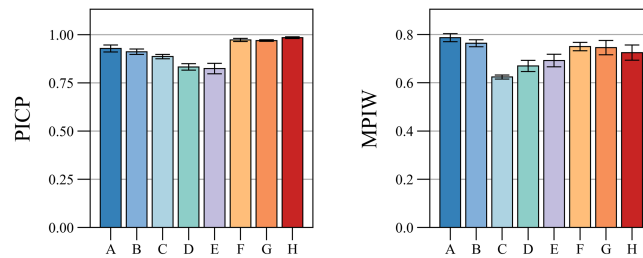


Figure 11: Prediction results of the comparative experiment: (a) MCA-TCN; (b) TCLSTM; (c) Attention-BiLSTM; (d) Transformer-LSTM; (e) CNN-Transformer; (f) Informer; (g) PatchTST; (h) Proposed CTCTN.

Table 7: Prediction results of the comparative experiment.

Method	RMSE	MAE	PICP	MPIW
MCA-TCN	0.2058 ± 0.0189	0.1654 ± 0.0175	0.9417 ± 0.0180	0.7950 ± 0.0164
TCLSTM	0.2057 ± 0.0086	0.1571 ± 0.0066	0.9049 ± 0.0141	0.7518 ± 0.0143
Attention-BiLSTM	0.1497 ± 0.0077	0.1117 ± 0.0066	0.9265 ± 0.0112	0.6356 ± 0.0082
Transformer-LSTM	0.2598 ± 0.0105	0.1933 ± 0.0093	0.8053 ± 0.0167	0.6702 ± 0.0231
CNN-Transformer	0.2732 ± 0.0075	0.2174 ± 0.0080	0.8406 ± 0.0271	0.7318 ± 0.0262
Informer	0.1525 ± 0.0091	0.1189 ± 0.0077	0.9732 ± 0.0080	0.7501 ± 0.0172
PatchTST	0.1591 ± 0.0072	0.1227 ± 0.0064	0.9697 ± 0.0042	0.7457 ± 0.0295
Proposed CTCTN	0.1236 ± 0.0087	0.0930 ± 0.0074	0.9829 ± 0.0046	0.6879 ± 0.0316

4.6 Analysis of Computational Complexity

For resource-constrained satellite applications, computational complexity of the prognostic method is critical. Table 8 presents the parameter counts, FLOP counts, and average inference time per sample for each comparative method. It can be observed that the proposed CTCTN achieves a computational efficiency that satisfies the requirements for real-time processing. Specifically, compared with Informer and PatchTST, the proposed CTCTN achieves better predictive performance while having lower inference time and fewer parameters.

Table 8: Computational complexity of all methods.

Method	Parameters	FLOPs	Inference Time (ms/sample)
MCA-TCN	19.786K	1.213G	0.3227
TCLSTM	3.826K	15.068M	0.0420
Attention-BiLSTM	11.810K	665.662M	0.0841
Transformer-LSTM	8.050K	130.057M	0.1236
CNN-Transformer	3.690K	25.412M	0.3562
Informer	113.778K	5.585G	3.112
PatchTST	102.194K	803.718M	1.0177
Proposed CTCTN	69.970K	184.490M	0.1463

4.7 Discussion

Based on the experiments presented above, the proposed CTCTN achieves superior prognostic performance compared with existing methods in the satellite SAP degradation scenario. This advantage arises from the alignment between the network architecture and the multi-scale characteristics of the degradation process. Traditional TCN-based methods, although capable of enlarging the receptive field through stacked DCCs, still process sequences using local sliding windows, which inherently limits their ability to model global dependencies. Attention-based methods excel at capturing long-range dependencies but exhibit limited capability in modeling local temporal features. CTCTN addresses these limitations through a cascaded design that leverages complementary strengths. The front-end DS-TC block extracts local features while simultaneously reducing sequence length. The subsequent Transformer module then performs multi-head self-attention within the compressed feature space, enabling each time step to capture long-term dependencies across the entire degradation process. Fig. 12 shows the heat maps of the attention weights of the 4 heads in the Transformer module. The x-axis and y-axis represent the time steps being focused on and the current time step, respectively. From the diagonal distribution to the global distribution, the different distribution patterns of the attention weights indicate that features of the different time scales have been learned by the Transformer module. This progressive fusion strategy makes CTCTN superior to both single-module architectures and traditional hybrid structures. This cascaded design constitutes the core innovation of this paper.

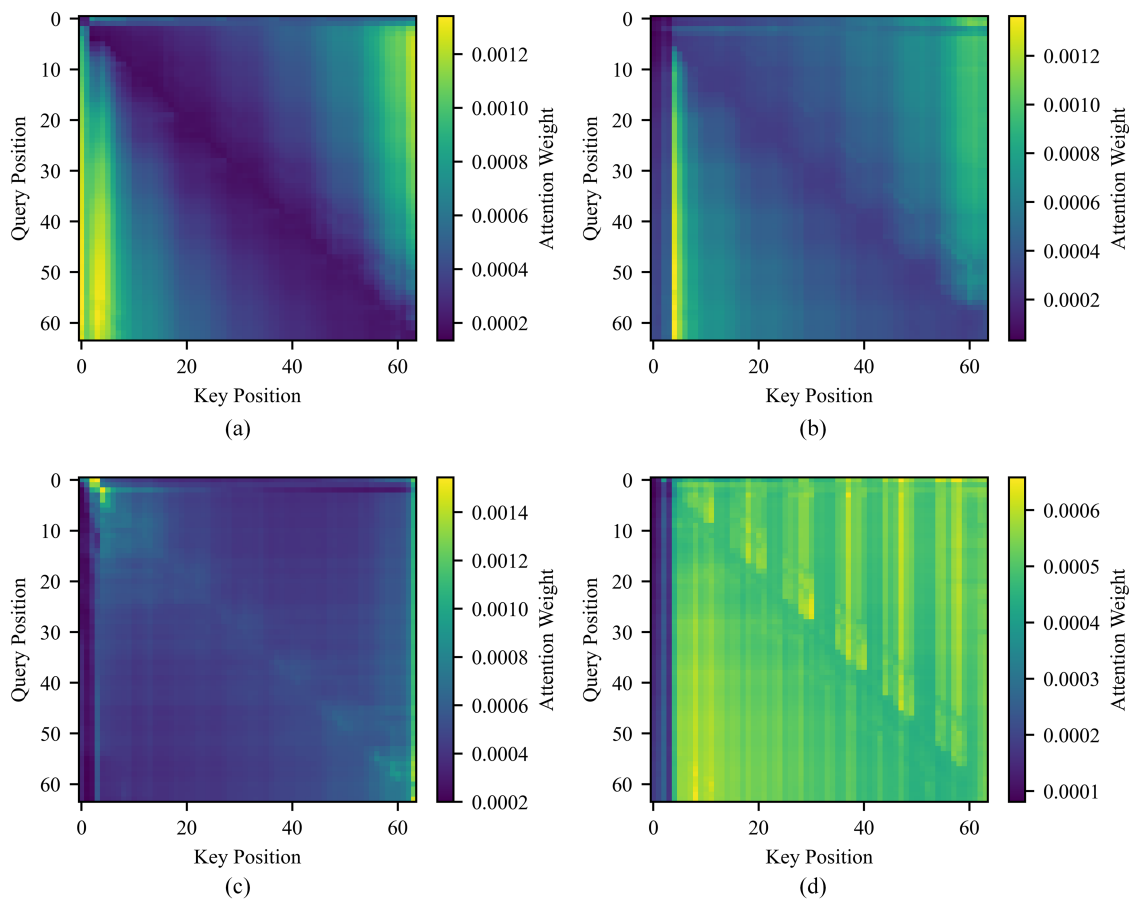


Figure 12: Heat maps of attention weights from Transformer: (a) Head 2; (b) Head 5; (c) Head 8; (d) Head 15.

While CTCTN demonstrates robust prognostic capability in simulated scenarios, it presents several challenges against transferring CTCTN to other prognostics domains, which also indicate promising directions for future research. The distribution differences between the source domain and the target domain limit the generalizability of the prediction methods. Therefore, cross-domain research represents a potential direction.

5 Conclusions

To address the challenges of failure prognosis and uncertainty quantification for SAP degradation in LEO satellite, this paper proposes CTCTN. Firstly, an AAP is used to align feature dimensions between the DS-TC blocks and the Transformer modules, enabling effective extraction of multi-scale degradation features. Secondly, during training, a heteroscedastic Huber loss function is designed to jointly optimize the dual output branches of the CTCTN for predicting the RUL mean and variance. Finally, during the testing stage, the MCD method is used to quantify both epistemic and aleatoric uncertainty and estimates the predictive confidence intervals, effectively capturing uncertainty arising from multi-scale features. A satellite model is developed, and a run-to-failure dataset is constructed to validate the proposed CTCTN method under the SAP performance degradation scenario of LEO satellite. Experimental results demonstrate that the proposed CTCTN method not only achieves more accurate RUL predictions but also effectively quantifies the uncertainty associated with multi-scale degradation features.

Acknowledgement: The authors were partially supported by the Key Laboratory of Spacecraft Design Optimization and Dynamic Simulation Technologies, Ministry of Education.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Yunfeng Dong; methodology, Yu Shi; software, Yu Shi; validation, Yu Shi and Lu Tian; formal analysis, Yu Shi and Lu Tian; investigation, Yu Shi; resources, Yu Shi; data curation, Yu Shi; writing—original draft preparation, Yu Shi and Yunfeng Dong; writing—review and editing, Yu Shi, Yunfeng Dong and Lu Tian; visualization, Yu Shi; supervision, Yunfeng Dong; project administration, Yunfeng Dong. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AAP	Adaptive Average Pooling
CBAM	Convolutional Block Attention Module
CI	Confidence Interval
CLS	Classification
CMGs	Control Moment Gyros
CNN	Convolutional Neural Network
CTCTN	Cascade Temporal Convolution and Transformer Network
DCC	Dilated Causal Convolution
DS-TC	Depthwise Separable Temporal Convolution
EMD	Empirical Mode Decomposition
EOL	End of Life

FPT	First Prediction Time
GAP	Global Average Pooling
GELU	Gaussian Error Linear Unit
HI	Health Indicators
ISHM	Integrated System Health Management
LEO	Low Earth Orbit
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MCD	Monte Carlo Dropout
MHAP	Multi-Head Attention Pooling
MPIW	Mean Prediction Interval Width
PICP	Prediction Interval Coverage Probability
ReLU	Rectified Linear Unit
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RUL	Remaining Useful Life
SAP	Solar Array Paddle
SVM	Support Vector Machine
TCN	Temporal Convolutional Network

References

1. Ranasinghe K, Sabatini R, Gardi A, Bijjahalli S, Kapoor R, Fahey T, et al. Advances in integrated system health management for mission-essential and safety-critical aerospace applications. *Prog Aerosp Sci.* 2022;128(2):100758. doi:10.1016/j.paerosci.2021.100758.
2. Foster MA. Practical system to remove lethal untracked orbital debris. *J Aerosp Inf Syst.* 2022;19(10):661–7. doi:10.2514/1.i1010985.
3. Gambi JM, Phipps C, Garcia del Pino ML, Mosser J, Weinmüller EB, Alderete M. Deflection of dangerous middle-size LEO debris with autonomous space-based laser brooms via surgical actions. *Acta Astronaut.* 2024;217(4):75–88. doi:10.1016/j.actaastro.2024.01.021.
4. Žilková D, Šilha J, Vojtek P, Rodriguez-Villamizar J, de Leon J, Matlovič P, et al. Connecting laboratory and spectroscopic observations of aerospace materials to characterize the reflectivity of artificial space objects and debris in LEO regimes. *Acta Astronaut.* 2025;236(9):479–86. doi:10.1016/j.actaastro.2025.07.003.
5. Huang JG, Han JW, Li HW, Cai MH, Li XY. Investigation on the surface damage to solar cells by impacts of space micro-debris on low earth orbit. *Acta Phys Sin.* 2008;57(12):7950. doi:10.7498/aps.57.7950.
6. Jiang D, Zhang P, Zhang Y. The study of space debris and meteoroid impact effects on spacecraft solar array. In: Kleiman J, editor. *Protection of materials and structures from the space environment.* Berlin/Heidelberg, Germany: Springer; 2017. p. 337–45. doi:10.1007/978-3-319-19309-0_34.
7. Liu X, Zhang D, Li A, Sun F, Yang Y. Study on the influence of solar array damage on satellite power system. In: *Proceedings of the 2018 10th International Conference on Modelling, Identification and Control (ICMIC); 2018 Jul 2–4; Guiyang, China.* doi:10.1109/ICMIC.2018.8529891.
8. Meng Y, Zhang D, Wang C, Liu Z, Zhu L, Li A. Modeling and analysis of non-linear phenomena of satellite power system in space environment and hazard-risk evaluations. *Electronics.* 2022;11(11):1756. doi:10.3390/electronics11111756.
9. Hedayati M, Barzegar A, Rahimi A. Fault diagnosis and prognosis of satellites and unmanned aerial vehicles: a review. *Appl Sci.* 2024;14(20):9487. doi:10.3390/app14209487.
10. Muthusamy V, Kumar KD. Failure prognosis and remaining useful life prediction of control moment gyroscopes onboard satellites. *Adv Space Res.* 2022;69(1):718–26. doi:10.1016/j.asr.2021.09.016.

11. Park HJ, Kim S, Lee J, Kim NH, Choi JH. System-level prognostics approach for failure prediction of reaction wheel motor in satellites. *Adv Space Res.* 2023;71(6):2691–701. doi:10.1016/j.asr.2022.11.028.
12. An D, Kim NH, Choi JH. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliab Eng Syst Saf.* 2015;133(7):223–36. doi:10.1016/j.ress.2014.09.014.
13. Zhao Z, Liang B, Wang X, Lu W. Remaining useful life prediction of aircraft engine based on degradation pattern learning. *Reliab Eng Syst Saf.* 2017;164(2):74–83. doi:10.1016/j.ress.2017.02.007.
14. Bienefeld C, Kirchner E, Vogt A, Kacmar M. On the importance of temporal information for remaining useful life prediction of rolling bearings using a random forest regressor. *Lubricants.* 2022;10(4):67. doi:10.3390/lubricants10040067.
15. Kumar RS, Singh AR, Narayana PL, Chandrika VS, Bajaj M, Zaitsev I. Hybrid machine learning framework for predictive maintenance and anomaly detection in lithium-ion batteries using enhanced random forest. *Sci Rep.* 2025;15(1):6243. doi:10.1038/s41598-025-90810-w.
16. Li Y, Zou C, Berecibar M, Nanini-Maury E, Chan JCW, van den Bossche P, et al. Random forest regression for online capacity estimation of lithium-ion batteries. *Appl Energy.* 2018;232:197–210. doi:10.1016/j.apenergy.2018.09.182.
17. Cheng Y, Zerhouni N, Lu C. A hybrid remaining useful life prognostic method for proton exchange membrane fuel cell. *Int J Hydrogen Energy.* 2018;43(27):12314–27. doi:10.1016/j.ijhydene.2018.04.160.
18. Nagulapati VM, Lee H, Jung D, Paramanatham S, Brigljevic S, Choi B, et al. A novel combined multi-battery dataset based approach for enhanced prediction accuracy of data driven prognostic models in capacity estimation of lithium ion batteries. *Energy AI.* 2021;5(11–12):100089. doi:10.1016/j.egyai.2021.100089.
19. Yan M, Wang X, Wang B, Chang M, Muhammad I. Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model. *ISA Trans.* 2020;98:471–82. doi:10.1016/j.isatra.2019.08.058.
20. Khelif R, Chebel-Morello B, Malinowski S, Laajili E, Fnaiech F, Zerhouni N. Direct remaining useful life estimation based on support vector regression. *IEEE Trans Ind Electron.* 2017;64(3):2276–85. doi:10.1109/TIE.2016.2623260.
21. Souto M, das Chagas Moura M, Lins I. Particle swarm-optimized support vector machines and pre-processing techniques for remaining useful life estimation of bearings. *Eksplot I Niezawodn.* 2019;21(4):610–8. doi:10.17531/ein.2019.4.10.
22. Song L, Lin T, Jin Y, Zhao S, Li Y, Wang H. Advancements in bearing remaining useful life prediction methods: a comprehensive review. *Meas Sci Technol.* 2024;35(9):092003. doi:10.1088/1361-6501/ad5223.
23. Xu W, Mao R, Han P, Yuan N, Li Y, Guo Y, et al. A comprehensive review of lithium-ion battery remaining useful life prediction: methodologies, datasets, performance metrics, and future perspectives. *Meas Sci Technol.* 2025;36(8):082001. doi:10.1088/1361-6501/adfb97.
24. Wang B, Lei Y, Yan T, Li N, Guo L. Recurrent convolutional neural network: a new framework for remaining useful life prediction of machinery. *Neurocomputing.* 2020;379:117–29. doi:10.1016/j.neucom.2019.10.064.
25. Guo L, Lei Y, Li N, Yan T, Li N. Machinery health indicator construction based on convolutional neural networks considering trend burr. *Neurocomputing.* 2018;292(4):142–50. doi:10.1016/j.neucom.2018.02.083.
26. Che C, Wang H, Fu Q, Ni X. Combining multiple deep learning algorithms for prognostic and health management of aircraft. *Aerosp Sci Technol.* 2019;94:105423. doi:10.1016/j.ast.2019.105423.
27. Wu Q, Ding K, Huang B. Approach for fault prognosis using recurrent neural network. *J Intell Manuf.* 2020;31(7):1621–33. doi:10.1007/s10845-018-1428-5.
28. Sirajul Islam M, Rahimi A. Fault prognosis of satellite reaction wheels using a two-step LSTM network. In: *Proceedings of the 2021 IEEE International Conference on Prognostics and Health Management (ICPHM); 2021 Jun 7–9; Detroit (Romulus), MI, USA.* doi:10.1109/icphm51084.2021.9486655.
29. Isbilen F, Bektas O, Konar M. Deep learning and similarity-based models for predicting turbofan engine remaining useful life: insights from the CMAPSS dataset. *Aeronaut J.* 2025;129(1337):2004–35. doi:10.1017/aer.2025.25.
30. van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, et al. WaveNet: a generative model for raw audio. *arXiv:1609.03499.* 2016.
31. Chen Q, Liu YB, Ge MF, Liu J, Wang L. A novel Bayesian-optimization-based adversarial TCN for RUL prediction of bearings. *IEEE Sens J.* 2022;22(21):20968–77. doi:10.1109/JSEN.2022.3209894.

32. Deng F, Bi Y, Liu Y, Yang S. Remaining useful life prediction of machinery: a new multiscale temporal convolutional network framework. *IEEE Trans Instrum Meas.* 2022;71:2516913. doi:10.1109/TIM.2022.3200093.
33. Hsu CY, Lu YW, Yan JH. Temporal convolution-based long-short term memory network with attention mechanism for remaining useful life prediction. *IEEE Trans Semicond Manuf.* 2022;35(2):220–8. doi:10.1109/TSM.2022.3164578.
34. Liu Q, Dai Z, Lai H, Chen M, Huang H, Fu J, et al. A novel RUL prediction method for rolling bearing: TcLstmNet-CBAM. *Sci Rep.* 2025;15(1):14055. doi:10.1038/s41598-025-98845-9.
35. Zou J, Lin P. Multichannel attention-based TCN-GRU network for remaining useful life prediction of aero-engines. *Energies.* 2025;18(8):1899. doi:10.3390/en18081899.
36. Dintén R, Zorrilla M, Veloso B, Gama J. Building of transformer-based RUL predictors supported by explainability techniques: application on real industrial datasets. *Inf Fusion.* 2026;127(2):103892. doi:10.1016/j.inffus.2025.103892.
37. Han F, Mo B. RUL prediction method based on sequential health index evaluation with multidimensional coupled degradation data. *PLoS One.* 2026;21(1):e0340645. doi:10.1371/journal.pone.0340645.
38. Xie S, Li Z, Dong Y, Chen P. Multigranularity batch sequential design method for component-level satellite design optimization. *IEEE Trans Aerosp Electron Syst.* 2024;60(6):8779–90. doi:10.1109/TAES.2024.3433325.
39. Toyota H, Nakamura T, Kanaya S, Sumita T, Hirai T, Kobayashi M. Evaluation of hypervelocity impact of micrometeoroids and orbital debris on next-generation space solar cells. *Jpn J Appl Phys.* 2023;62:SK1047. doi:10.35848/1347-4065/acd18b.
40. Yuan Y, Zhang J, Yang K, Li L, Wu H. Collision risk assessment for constellation satellites based on a space debris environment topological network model. *Chin J Aeronaut.* 2026;39(2):103762. doi:10.1016/j.cja.2025.103762.
41. Zheng JD, Zhou J, Pi XL, Zou C, Li YF, Xu KB, et al. Hypervelocity impact on volt-ampere characteristic of solar arrays by using two-stage light gas gun. *Acta Phys Sin.* 2021;70(18):188801. doi:10.7498/aps.70.20210458.
42. Li N, Lei Y, Lin J, Ding SX. An improved exponential model for predicting remaining useful life of rolling element bearings. *IEEE Trans Ind Electron.* 2015;62(12):7762–73. doi:10.1109/TIE.2015.2455055.
43. Nemani VP, Lu H, Thelen A, Hu C, Zimmerman AT. Ensembles of probabilistic LSTM predictors and correctors for bearing prognostics using industrial standards. *Neurocomputing.* 2022;491(5):575–96. doi:10.1016/j.neucom.2021.12.035.
44. Wang W, Lei Y, Yan T, Li N, Nandi A. Residual convolution long short-term memory network for machines remaining useful life prediction and uncertainty quantification. *J Dyn Monit Diagn.* 2022;1(1):2–8. doi:10.37965/jdmd.v2i2.43.
45. Kumar A, Parkash C, Kundu P, Tang H, Xiang J. Enhanced deep learning framework for accurate near-failure RUL prediction of bearings in varying operating conditions. *Adv Eng Inform.* 2025;65:103231. doi:10.1016/j.aei.2025.103231.
46. Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, et al. A review of uncertainty quantification in deep learning: techniques, applications and challenges. *Inf Fusion.* 2021;76(1):243–97. doi:10.1016/j.inffus.2021.05.008.
47. Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? *arXiv:1703.04977.* 2017.
48. Li G, Yang L, Lee CG, Wang X, Rong M. A Bayesian deep learning RUL framework integrating epistemic and aleatoric uncertainties. *IEEE Trans Ind Electron.* 2021;68(9):8829–41. doi:10.1109/TIE.2020.3009593.
49. Li Z, Shen S, Ye Y, Cai Z, Zhen A. An interpretable online prediction method for remaining useful life of lithium-ion batteries. *Sci Rep.* 2024;14(1):12541. doi:10.1038/s41598-024-63160-2.
50. Dong Y, Li Z, Lei M. Research on concept of digital satellite. *Aerosp Shanghai.* 2021;38:1–12. doi:10.19328/j.cnki.1006-1630.2021.01.001.
51. Ren M, Dong Y, Li C. Code generation technology of digital satellite. In: *Proceedings of ELM-2015 Volume 1.* Berlin/Heidelberg, Germany: Springer; 2016. p. 511–9. doi:10.1007/978-3-319-28397-5_40.
52. Mebarki N, Mouss LH, Bentrucia T, Benmoussa S. An integrated physical model and extant data based approach for fault diagnosis and failure prognosis: application to a photovoltaic module. *Microelectron Reliab.* 2025;168(5):115711. doi:10.1016/j.microrel.2025.115711.

53. Sgorlon Gaiatto C, Antonello F, Segneri D, Sousa B, Abascal Palacios B, Schiavo A, et al. A novel physics-based computational framework to model spacecraft solar array power under degradation: application to European space agency (ESA) cluster mission. *Acta Astronaut.* 2025;226(1):341–8. doi:10.1016/j.actaastro.2024.10.052.
54. Xie S, Dong Y, Liang Z. Genetic programming method for satellite optimization design with quantification of multi-granularity model uncertainty. *Aerosp Sci Technol.* 2025;156(3):109764. doi:10.1016/j.ast.2024.109764.
55. Lei M, Dong Y. Multi-granularity modeling method for effectiveness evaluation of remote sensing satellites. *Remote Sens.* 2023;15(17):4335. doi:10.3390/rs15174335.
56. Zhu B, Dong E, Cheng Z, Jiang K, Guo C, Yue S. An integrated attention-BiLSTM approach for probabilistic remaining useful life prediction. *Comput Mater Contin.* 2026;87(1):1–10. doi:10.32604/cmc.2025.074009.
57. Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: beyond efficient transformer for long sequence time-series forecasting. *Proc AAAI Conf Artif Intell.* 2021;35(12):11106–15. doi:10.1609/aaai.v35i12.17325.
58. Nie Y, Nguyen NH, Sinthong P, Kalagnanam JA. Time series is worth 64 words: long-term forecasting with transformers. *arXiv:2211.14730.* 2023.