This paper proposes a comprehensive framework for predictive maintenance of complex systems, such as aircraft turbofan engines, by integrating data-driven probabilistic remaining useful life (RUL) prognostics with maintenance planning for both single-component and multi-component systems.

The paper addresses the limitations of existing studies that focus either on point RUL estimates or on maintenance planning models that rely on generic probability distributions.

The proposed method, described in Section 2, utilizes a Convolutional Neural Network (**1**) architecture combined with Monte Carlo dropout (**2**) to estimate the probability density function of the RUL. The CNN is trained on the C-MAPSS dataset, which contains sensor data from turbofan engines. The architecture and hyperparameters of the CNN are detailed in subsections 2.2 and 3.1.

The performance of the probabilistic RUL prognostics is evaluated in Section 3 using metrics such as root mean square error (**3**) for mean RUL predictions, and α-coverage (**4**) and reliability diagrams for the reliability of the probabilistic estimates. The results show that the prognostics are reliable but have uncertainty (subsection 3.3).

Section 4 presents methods for integrating the probabilistic RUL prognostics into maintenance planning. For single-component systems, a renewal-reward process is formulated to determine the optimal replacement time (subsection 4.1). For multi-component systems, an integer linear programming (**5**) model is proposed, considering resource constraints such as maintenance slot availability and capacity (subsection 4.2).

The application of the proposed maintenance planning methodology to the C-MAPSS dataset is demonstrated in Section 5. Results for both single-engine and multi-engine maintenance planning are presented, highlighting the benefits of using probabilistic RUL prognostics (subsections 5.2 and 5.3). A long-term performance comparison with time-based maintenance and perfect RUL prognostics strategies is conducted using a Monte Carlo simulation, demonstrating the effectiveness of the proposed approach in reducing costs and failures (subsection 5.4).

In conclusion, the paper presents an end-to-end framework for data-driven predictive maintenance that integrates probabilistic RUL prognostics with maintenance planning. The case study on turbofan engines shows that the approach leads to reliable RUL prognostics, significant cost savings, and failure reductions compared to time-based maintenance, with performance close to the perfect prognostics scenario.

DEFINITIONS

1. CNN
   1. Deep learning algorithm that uses convolutional layers to automatically learn hierarchical features from input data, making it particularly effective for image and signal processing tasks.
   2. A convolutional layer applies a set of learnable filters (kernels) to the input data, performing convolution operations to extract local patterns and features while preserving spatial relationships in the data.
      1. Each filter is a small matrix of numbers that looks for specific patterns or features in the input.
      2. The filter slides over the input data, performing a mathematical operation called convolution.
      3. Convolution multiplies the values in the filter with the corresponding values in the input and sums them up.
      4. This process is repeated for each position of the filter on the input, creating a new matrix called a feature map.
      5. The feature map highlights the areas in the input where the filter pattern is found.
      6. By applying many different filters, the convolutional layer can detect various patterns and features in the input data.
2. Monte Carlo dropout
   1. Technique used in deep learning where dropout (randomly setting a fraction of input units to zero) is applied during both training and inference to approximate Bayesian inference, allowing the estimation of uncertainty in the model's predictions.
      1. Bayesian inference is a method of statistical inference in which **Bayes' theorem** is used to update the probability for a hypothesis as more evidence or information becomes available



1. RSME
   1. Widely used metric for evaluating the accuracy of a regression model by measuring the average magnitude of the errors between the predicted and actual values, calculated as the square root of the mean of the squared differences between predictions and true values.
2. α-coverage
   1. α-coverage is a metric used to evaluate the reliability of probabilistic predictions, such as those generated by machine learning models. It measures the proportion of true values that fall within a specified confidence interval (defined by the parameter α) of the predicted probability distribution, with higher α-coverage indicating more reliable predictions.
3. ILP
   1. mathematical optimization technique used to find the best solution to a linear objective function subject to linear constraints, where some or all of the decision variables are restricted to integer values, making it suitable for solving complex problems with discrete or combinatorial aspects.