Forecasting the Global Demand of HIV medicines using Machine Learning Techniques

Rishabh Verma, *Student Member, IEEE,* Prof. Gyan Prakash, *Member, IEEE,*

***Abstract*—In this research paper, we explore the application of machine learning techniques to forecast the global demand for HIV medicines. Leveraging the power of data-driven approaches, we aim to address the challenges associated with accurately predicting demand patterns in the pharmaceutical supply chain. By collecting and analyzing data on HIV medicine consumption, we develop a novel framework that integrates various machine learning algorithms. Through supervised learning methods, we train models to forecast future demand trends, while also incorporating these techniques to optimize inventory management and allocation policies. This research integrates advanced preprocessing methodologies such as data normalization and feature selection to refine the data and identify significant predictors. The model architecture encompasses of supervised learning algorithms, including decision trees, Extra Trees and neural networks, facilitating precise prediction of future demand trends while optimizing allocation policies. Through extensive evaluation using real-world data, this research demonstrates superior performance compared to traditional methods, showcasing its potential to enhance the efficiency of HIV medicine distribution on a global scale.**

***Index Terms*—HIV, Human Immunodeficiency Viruses, Decision Trees, Extra Trees, Supervised Learning Algorithms,**

I. Introduction

IN today's age of industrial advancement, our society relies heavily on a vast array of electrical and mechanical systems, integral to sectors ranging from agriculture and power supply to transportation. Regardless of their complexity or scale, all systems necessitate maintenance operations at specific intervals. In the industrial landscape, maintenance techniques fall into three primary categories: reactive maintenance, preventive maintenance, and predictive maintenance. Reactive maintenance entails repairing machinery only after a failure occurs, leading to costly downtimes and emergency repairs. Preventive maintenance operates proactively, aiming to prevent failures through scheduled, time-based interventions. In contrast, predictive maintenance (PdM) adopts a proactive approach, estimating an equipment's time-to-failure by analyzing various machine parameters obtained through condition-based monitoring (CBM). PdM focuses on forecasting errors by modeling degradation trends between input sensors and the machine's time-to-failure duration. This approach offers significant advantages, eliminating unplanned downtime, reducing maintenance costs, and maximizing machine lifetime, particularly in safety-critical scenarios. In various industries, plants entail fixed costs such as equipment, land, and wages, along with variable expenses like power and raw materials. The profitability of an organization hinges on generating substantial profit after selling manufactured products in the market. When equipment or components fail, not only does this lead to a loss of profits, but it also incurs additional unplanned maintenance costs. During downtime, fixed costs accumulate without corresponding production, while variable costs rise due to increased maintenance expenses and decreased consumables costs. These losses persist until the plant resumes regular operations. Severe equipment failures can result in downtime costs far exceeding profits generated within the same timeframe. In some cases, equipment is replaced prematurely, leading to the underutilization of its useful lifespan. Alternatively, equipment failures can cause unexpected downtimes, further impacting productivity and profitability. Accurate estimation of the equipment's useful life is crucial for cost-effective operations. The identification of potential failure and function failure based on degradation symptoms is a crucial step in understanding equipment lifespan. Prognostics and Health Management (PHM) are pivotal aspects of managing machinery and processes, with Remaining Useful Life (RUL) estimation at the core [1]. RUL prediction provides valuable insights into the system's current health status, indicating performance degradation and offering prevention strategies against abrupt failures. RUL estimation, a critical need in today's economic climate, finds applications in essential components like aircraft engines and nuclear power plants. Conventional methods of estimating useful life typically consider static machine conditions. However, in the era of Industry 4.0, dynamic systems can be monitored in real-time, enabling precise RUL estimations.

Market forecasts indicate a significant growth in predictive and prescriptive maintenance, with an estimated market value of $22.72 billion by 2026 and a Compound Annual Growth Rate (CAGR) of 19.68% [2]. Machine downtime incurs substantial costs, with averages reaching $260,000 per hour across various sectors. In specific industries like automotive, downtime costs escalate rapidly, emphasizing the financial significance of effective maintenance strategies. Surprisingly, a substantial portion of industrial sectors lacks awareness regarding the timing of necessary equipment maintenance or replacement due to a lack of RUL estimation knowledge. In manufacturing, up to 20% of machine downtime results from cutting tool failures, underscoring the need for accurate maintenance planning. Effective system monitoring not only enhances productivity by 10-40% but also leads to cost savings of up to 40%. These statistics emphasize the vital role of precise RUL estimations and efficient maintenance strategies in optimizing industrial productivity and minimizing costs. A prominent illustration of the criticality of predictive maintenance is evident in the aviation industry, where continuous monitoring of aircraft engine performance is paramount [3], [4], [5]. Fault diagnostics and prognostics for aircraft engines have garnered substantial attention, with a key focus on accurately determining the engine's Remaining Useful Life (RUL) [6], [7]. RUL prediction models, rooted in degradation trends observed among various condition monitoring sensors, form the basis for strategic maintenance planning. These models enable targeted development of maintenance strategies, minimizing unplanned downtimes, and extending machine lifespans, especially in safety-critical contexts. Early anomaly detection and timely warnings of potential failures are essential for optimizing system utilization.

Predictive maintenance is crucial in the aviation industry, particularly for aircraft engines, focusing on accurate Remaining Useful Life (RUL) prediction. RUL models, derived from sensor data, enable strategic maintenance, minimize downtime, and enhance machine lifespans. Early anomaly detection is vital, with three main approaches: physical models, data-driven techniques, and hybrid methods, each offering unique advantages. Physical model-based methods [8], [9] involves a deep understanding of the machine's physical architecture, employing laws of physics to create mathematical models for RUL estimation. Data-driven techniques [10], [11] conversely, leverage statistical and machine learning algorithms to uncover patterns within sensor data, making them suitable for complex industrial machinery. Hybrid approaches [12] amalgamate the strengths of both physical and data-driven models.

In prognostics, two prominent methodologies have emerged: model-based and data-driven approaches. The former, as the name suggests, involves an intricate understanding of the physical machine architecture, incorporating the laws of physics to derive mathematical models for estimating Remaining Useful Life (RUL). However, these mathematical models often necessitate simplifying assumptions, introducing uncertainties in the context of complex industrial machinery. Consequently, the accuracy of RUL predictions can be compromised.

Data-driven prognostic techniques, employing diverse statistical and machine-learning algorithms, identify sensor data trends without requiring exhaustive machine understanding. Particularly effective for complex industrial machinery, these techniques have gained prominence.

In the past decade, data-driven prognostic methods have experienced a surge in popularity among researchers. These models estimate RUL by analyzing degradation trends and target trajectories within sensor data. Notably, deep learning techniques, including autoencoders, convolutional neural networks (CNN), long short-term memory (LSTM) networks, and their variants, have achieved significant success in various fields, such as computer vision, speech recognition, video segmentation, and predictive maintenance [13].

However, the efficacy of deep learning algorithms is contingent upon access to substantial data for offline training, posing a significant challenge in the context of prognostics, where obtaining runtime-to-failure sensor data, especially for new machines, proves arduous. While intentional system failure can be an option, it is a protracted, undesirable, and costly approach. Consequently, researchers have turned to publicly available datasets for evaluation purposes. In this study, the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset, a simulation of turbofan jet engine data provided by NASA's Prognostics Center of Excellence, was utilized. This dataset comprises multiple multivariate time series units, each featuring varying numbers of engines, each with distinct Remaining Useful Life (RUL) values. Additionally, the dataset encompasses twenty-one sensors, capturing three operating conditions concerning the engine's time cycle [14].

Regarding the specific focus on turbofan engines, recent research efforts have explored the application of deep learning methods, including CNN and LSTM networks, alongside their combinations and variations, for RUL prediction [15], [16]. LSTM networks have exhibited superior performance compared to CNN-based models. The strength of LSTM networks lies in their aptitude for handling time-series data, enabling them to grasp the temporal features inherent in multivariate systems and minimize the Root Mean Square Error (RMSE) concerning target predictions [17].

In this study, our research concentrates on enhancing the RUL prediction. We achieve this by implementing crucial preprocessing steps on the sensor data before feeding it into the SAFETNet network. These preprocessing steps include correlation analysis, data filtering, normalization, and the utilization of a modified piecewise linear degradation model for determining the starting point of degradation. It has been demonstrated that accurately determining the starting point of degradation, also known as the initial RUL, significantly impacts the precision of RUL predictions [18]. Through these meticulous steps, our study aims to empower SAFETNet networks to predict RUL with unparalleled accuracy, thereby contributing substantially to the field of prognostics.

In the realm of aerospace engineering, specifically concerning aircraft turbofan engines, the accurate prediction of Remaining Useful Life (RUL) remains a significant challenge. The complexity of these engines, combined with the high stakes involved in their operation, necessitates precise prognostic methods. Traditional maintenance approaches, relying on fixed schedules or reactive responses, are increasingly inadequate for these intricate systems. The problem at hand is how to leverage advanced machine learning techniques to develop a predictive maintenance model that can reliably estimate the RUL of aircraft turbofan engines. This prediction must be accurate, timely, and adaptable to diverse operational conditions. The motivation behind this research stems from the critical implications of accurate RUL prediction for aircraft turbofan engines. Unplanned downtime, premature failures, and inefficient maintenance practices can lead to substantial financial losses, compromise safety, and undermine operational efficiency. The aerospace industry, characterized by its stringent safety standards and operational demands, requires innovative solutions to ensure uninterrupted and reliable engine performance.

Additionally, the advent of Industry 4.0 and the increasing availability of vast datasets have opened new avenues for predictive maintenance. By developing robust predictive models, this research aims to revolutionize maintenance strategies in the aerospace sector. The ability to anticipate engine failures in advance reduces downtime and optimizes maintenance schedules, leading to cost savings and enhanced operational efficiency.

Moreover, the research community's constant pursuit of more accurate and efficient predictive maintenance techniques drives this study. By addressing the challenges related to data quality, algorithm complexity, and real-time applicability, this research project endeavors to contribute cutting-edge methodologies to the broader field of prognostics and predictive maintenance. The potential impact of this research extends beyond the aerospace industry, potentially influencing predictive maintenance practices in various sectors, making it a compelling and timely area of study.

The main contributions of this research work are as follows:

1) The primary contribution lies in the introduction of SAFETNet, a pioneering deep architecture for prognostic tasks. SAFETNet combines the capabilities of Self-Attention (SA), Temporal Convolution Network (TCN), and Squeeze & Excitation (SE) blocks. This unified framework is meticulously designed for fault prognosis, excelling under intricate time-varying operating conditions. The TCN enhanced with an attention mechanism, efficiently captures temporal feature representations from raw sensor data, enabling end-to-end Remaining Useful Life (RUL) prediction with an optimized training process.
2) This research also makes a significant contribution to the advancement of SAFETNet by incorporating an extensive data preprocessing approach for Remaining Useful Life (RUL) prediction. Emphasizing systematic data quality assessment and corrective measures, the study ensures the generation of high-quality input data for data-driven models. This meticulous preprocessing methodology enhances the reliability and accuracy of the RUL prediction process by addressing data inconsistencies and anomalies.

The structure of the subsequent sections in this research paper is outlined as follows:

Section 2 presents an in-depth review of the current methodologies employed in predicting the Remaining Useful Life (RUL) of turbofan engines. In Section 3, the paper elucidates the novel model proposed for RUL prediction, detailing its methodology. Section 4 delves into the dataset acquisition and preprocessing techniques, training procedures, and the performance evaluation of the proposed model. This section also includes a comparative analysis with other existing models. Finally, Section 5 encapsulates the findings, summarizes the research outcomes, and delineates avenues for future research.

II. Literature Review

In this study, our literature review has been conducted across three pivotal domains closely associated with the estimation of Remaining Useful Life (RUL) for turbofan engines: Statistical Knowledge-Based Models, Classical Degradation Models, and Data-Driven Models. Each of these domains represents a distinct approach in the existing body of research concerning RUL prediction. This comprehensive exploration enables us to critically analyze and integrate insights from diverse methodologies, ensuring a robust foundation for our proposed research framework.

A. Statistical knowledge-based model

The Statistical knowledge-based model mainly uses past equipment failure or breakdown data for statistical characterization and makes fault prediction. In the realm of turbofan engine Remaining Useful Life (RUL) estimation, a variety of techniques have been explored, showcasing the evolution of methodologies in predictive maintenance [19]. Traditional model-based approaches have traditionally relied on algorithms such as the Kalman filter (KF), extended Kalman filter (EKF), and particle filters. These algorithms have been instrumental in formulating mathematical representations of machines using multi-sensor time series sequence data [20], [21].

B. Classical Degradation Model

Classical degradation models, such as the Eyring model or Weibull distribution, have been implemented in past studies to predict the RUL of turbofan engines [22]. Salahshoor et al. [23] contributed to this field by introducing a unified framework of EKF-based design for sensor data fusion algorithms. This innovation significantly enhanced the detection and diagnosis of degradation trends and system faults. Another noteworthy study by Ordonez et al. [24] combined the power of the auto-regressive integrated moving average (ARIMA) model and support vector machine (SVM) methods. By integrating these techniques, they successfully estimated the RUL, demonstrating the potential of hybrid models in this domain.

C. Data Driven Models

Furthermore, researchers have delved into learning-oriented approaches, emphasizing the importance of leveraging prior knowledge about degradation models. Khelif et al. [25] presented an innovative SVR model capable of establishing a direct correlation between multivariate sensor data or health indices and the RUL of aircraft turbofan engines. This shift toward machine learning signifies a progressive trend in RUL estimation, marking a departure from traditional methodologies and embracing the power of data-driven models for more accurate and efficient predictions.

In the landscape of turbofan engine RUL prediction, deep neural network-based methods have emerged as a dominant force, leveraging advanced algorithms and innovative architectures to enhance accuracy and efficiency.

Zhang et al. [26] pioneered a multi-objective evolutionary algorithm, expanding deep belief networks into parallel networks to achieve diversity and accuracy in RUL prediction. These networks demonstrated remarkable precision, particularly in complex operational scenarios and amidst input data noise. Saeidi et al. [27] introduced a naive Bayesian classification algorithm, incorporating a moving average filter during pre-processing to eliminate noise from sensor data. The dataset was categorized based on time cycles, enabling urgent maintenance detection for cycles between 0 to 50. This method showcased the power of efficient data preprocessing in enhancing RUL predictions. Zheng et al. [28] proposed an LSTM network combined with a piecewise linear function for RUL estimation, achieving impressive results through the application of piecewise linear functions and data normalization. Wei et al. [29] introduced a Bi-directional LSTM network capable of learning high-level features bidirectionally. This novel approach utilized the backpropagation algorithm, enhancing the network's learning capacity. Wang et al. [30] devised a hybrid network for turbofan engines, leveraging LSTM networks to identify trends and hidden patterns in long sequence sensor data. Short-duration sequences were analyzed through a time window method coupled with gradient boosting regression (GBR). This two-stage approach, involving offline learning of degradation patterns with LSTM networks and online analysis with TW-GBR, proved effective, especially when combined with standardization and sensor selection criteria. In the realm of convolutional neural networks, Babu et al. [31] proposed a deep CNN regression network for RUL estimation. This model utilized two-dimensional convolutional layers for feature extraction, followed by a fully connected regression layer for precise predictions. Li et al. [32] introduced a deep convolutional neural network (CNN) tailored for RUL estimation. Their architecture facilitated feature extraction from prepared 2-D sensor data through convolution layers, flattened for input into a multilayer perceptron model with dropout layers, optimizing RUL predictions. Jayasinghe et al. [33] combined CNN-LSTM networks, employing data augmentation and normalization to avoid overfitting. A unique feature of this model was the use of temporal convolution, which acted as a bridge between the output of the 1-D convolution layer and the input of the LSTM layer. Hong et al. [34] stacked layers comprising 1-D convolution, residual, LSTM, and Bi-LSTM, performing correlation analysis on turbofan engine sensor data to enhance predictions. Mo et al. [35] introduced a multi-head neural network for RUL prediction, employing parallel branches of CNN layers alongside LSTM networks. They employed the Fisher method, recursive least squares, and single exponential smoothing for optimal performance, offering a novel approach in RUL estimation. Zhao et al. [36] proposed an adjacent neural network model leveraging the Markov property, predicting the next state of a sequence based solely on present states, showcasing the potential of innovative methodologies in this field. Additionally, the utilization of piecewise linear degradation models has gained traction in RUL prediction techniques. Lan et al. [37] proposed an LSTM algorithm employing a piecewise linear degradation model, dividing the dataset into time windows. Geometric distances between these windows determined the initial RUL or starting point of degradation. While the selection of window size and thresholds were not explicitly explained, this model served as inspiration for an improved version applied to the complete dataset, resulting in enhanced prediction accuracy. In the field of predictive maintenance, accurately predicting the Remaining Useful Life (RUL) of industrial machines under dynamic operating conditions is a pressing challenge. Liu et al. [38] have introduced the Dual Attention-Based Temporal Convolutional Network (DATCN), a novel deep learning method that integrates sensor data with complex operating condition information. DATCN employs dual attention modules within a Temporal Convolutional Network (TCN) framework, allowing it to adapt to diverse operating conditions and enhance RUL predictions. The study compares DATCN with various state-of-the-art methods, utilizing evaluation metrics like root mean square error (RMSE) and score metrics. Despite promising results, challenges related to negative transfer effects and domain adaptation persist, indicating the need for further exploration. Future research aims to refine transfer learning strategies for addressing prognostic challenges in varying operational contexts and multiple fault modes. The study also provides valuable visualizations, showcasing the learned attention weights and their influence on RUL predictions, enriching the insights into the proposed DATCN method's effectiveness.

The study by Sayyad et al. [39] explores proactive maintenance strategies for milling cutting tools by estimating Remaining Useful Life (RUL) through data-driven techniques, aiming to enable proactive maintenance and reduce unplanned downtime caused by tool failures. The authors utilized a systematic review process, employing Scopus, Web of Science, and IEEE databases to conduct a comprehensive literature survey. From 91 initial records, they narrowed down to 37 core documents relevant to milling RUL estimation. The study's limitations encompass challenges related to the disturbance of Acoustic Emission (AE) signals, sensitivity issues with motor current sensors, the high dimensionality of raw sensor data necessitating advanced processing techniques, and the accuracy limitations of direct sensors like optical microscopes. Additionally, data scarcity in transfer learning methods poses a hurdle, emphasizing the need for further exploration in self-supervised learning algorithms.

In this review, advanced predictive maintenance methods for industrial machinery are explored, with a focus on innovative approaches like deep adversarial learning in intelligent manufacturing systems. Ziu et al. [2] aims to predict machine states and faults in real time, enabling proactive maintenance and minimizing operational disruptions and costs. The study employs diverse datasets and intricate methodologies, integrating external sensor data and internal machine information. The LSTM-GAN algorithm is utilized, and the model's performance is evaluated through various metrics. The paper discusses limitations, suggests future research directions, and rigorously compares the proposed method with state-of-the-art algorithms. Overall, this review provides a comprehensive overview of cutting-edge predictive maintenance techniques, emphasizing the transformative potential of advanced machine learning in industrial maintenance strategies.

Essien [40] addressed the challenge of multistep time-series forecasting, focusing specifically on machine speed prediction in a manufacturing plant within the realm of smart manufacturing. They utilize a dataset from a bodymaker machine and employ the 2-DConvLSTMAE model for prediction, comparing its performance against baseline models. Despite its accuracy, the 2-DConvLSTMAE model demands longer training times, raising concerns about real-world efficiency. The study lacks transparency regarding hyperparameter optimization details and the broader applicability of the proposed model beyond the specific dataset and context. Future scopes encompass advanced deep learning for time-series forecasting, integration of artificial intelligence in manufacturing, traffic predictive analytics, sensor technologies, and enhancing human-machine interactions. The evaluation relies on metrics like RMSE, MAE, and sMAPE, although a comprehensive analysis of these metrics' suitability and comparison to other standard metrics is absent.

The extensive survey on predictive maintenance (PdM) of industrial equipment was done by Zhang et al. [14], they focus on the vital role of data-driven methods within the landscape of smart manufacturing and industrial big data. Their objective is to tackle the challenge of predicting equipment failures and optimizing maintenance, employing various sophisticated algorithms. They underscore critical limitations, including data validity concerns, high costs of data acquisition, and the interpretability issues inherent in advanced algorithms. Looking forward, the authors advocate for rigorous data validation processes, integration of advanced AI techniques, and the development of real-time predictive maintenance using big data frameworks. They emphasize the need for models not just to be accurate but also interpretable, pushing for a deeper understanding of the reasons behind equipment health conditions. Moreover, they highlight the importance of addressing practical challenges and encouraging diverse industrial applications. Although the survey primarily emphasizes accuracy as an evaluation metric, it provides a roadmap for researchers and practitioners, urging them to innovate and overcome challenges in the dynamic field of predictive maintenance in industrial contexts.

Malawade et al. [41] delve into the innovative use of Hierarchical Temporal Memory (HTM), a Neuroscience-Inspired Algorithm, for anomaly detection. Their approach involves stages like encoding, spatial pooling, temporal pooling, prediction, and anomaly detection within the HTM framework. The study addresses the crucial problem of predicting equipment failures before they occur, enabling efficient maintenance and cost reduction. The authors utilize the bearing dataset and a synthesized 3-D printer dataset for experimentation. While the HTM methodology demonstrates noise robustness and data efficiency compared to traditional techniques like Long Short-Term Memory (LSTM), limitations include the need for more evaluation on real-world examples and further exploration in Remaining Useful Life (RUL) estimation. The study also highlights HTM's advantages in terms of reduced training time and its ability to perform well without extensive domain-specific knowledge. The document concludes with a comparison of HTM against various state-of-the-art methods, emphasizing HTM's effectiveness in predictive maintenance scenarios.

The challenges of asset management in Industry 4.0 were tackled by Teoh et al. [42]. The researchers aim to develop an efficient resource management technique. Focused on predictive maintenance for manufacturing equipment, they propose a Genetic Algorithm (GA) based resource scheduling solution. The study utilizes original and amended datasets, employing GA for asset management and logistic regression for predictive maintenance. The proposed technique is compared with traditional scheduling methods such as MinMin, MaxMin, FCFS, and Round Robin using metrics like execution time, cost, and energy usage. While specific limitations and future scopes are not provided, the study emphasizes the importance of advancements in cloud computing, fog computing applications, IoT technologies, and energy-efficient computing in the context of Industry 4.0. The evaluation metrics used include execution time, cost, and energy usage, providing a comprehensive assessment of the proposed GA-based resource management technique against traditional algorithms.

Ong et al. [43] address the challenge of optimizing resource allocation and maintenance task scheduling in manufacturing systems to maximize overall production uptime. Researchers aim to integrate predictive maintenance (PdM) and manpower resource management, usually treated independently, to enhance production throughput and energy consumption. Although the specific dataset and evaluation metrics are not provided, the study explores frameworks such as genetic algorithms, deep reinforcement learning (DRL), and knowledge transfer (TL) methods. They propose a two-stage framework integrating TL and DRL models into a generic PdM-based resource management approach. The study identifies several limitations, including real-time scheduling complexities, suboptimal performance, distribution shifts, and scarcity of expert demonstration data. Future scopes include studying the impact of demonstration data size on TL performance, augmenting data sources, optimizing learning efficiency, exploring TL scalability, and investigating the robustness of TL techniques in various scenarios. The document does not specify the exact evaluation metrics used or the comparison with state-of-the-art methods.

After a thorough literature review spanning Statistical Knowledge-Based Models, Classical Degradation Models, and Data-Driven Models, a noticeable research gap emerges.

- Existing models, such as LSTM and Bidirectional LSTM, confront challenges in effectively capturing long-term dependencies within degradation sequences. Future research should focus on addressing issues like vanishing gradients to enhance the accurate representation of intricate temporal patterns, thereby improving Remaining Useful Life (RUL) predictions.
- CNNs, primarily designed for spatial feature extraction, might face limitations in capturing nuanced temporal dynamics crucial for RUL prediction. Future research endeavors should explore innovative approaches to adapt CNNs for better modeling of sequential dependencies, closing the gap in accurately representing the evolving health conditions of industrial machinery over time.
- The integration of data from diverse sensors is crucial for accurate Remaining Useful Life (RUL) estimation. However, existing challenges related to varied formats, sizes, and measurement units hinder seamless integration. Future research should concentrate on developing AI-based frameworks adept at handling multi-sensor data, focusing on data standardization, normalization, and fu-

sion techniques to ensure a common analysis framework.

- Environmental factors, such as noise and temperature variations, introduce uncertainties into sensor data, impacting the precision of AI-based RUL predictions. Addressing this gap requires effective data preprocessing techniques to filter out noise and enhance input signal quality. Furthermore, there is a need for the development of autocorrecting AI algorithms tailored to noisy industrial environments, accompanied by validation metrics ensuring accurate RUL estimations despite environmental disturbances.

III. Problem Formulation

This study aims to investigate the fault prognosis problem in the context of dynamic operating conditions, where multiple industrial equipment operate within complex environments. The dataset utilized for this investigation comprises Condition Monitoring (CM) data, which includes multisensor inputs and specific operational conditions. This diverse dataset gives rise to two distinct domains: the source domain, which consists of labeled training data, and the target domain, comprising unlabeled testing data. Addressing the distribution gap between these domains is pivotal for effective model generalization. To accomplish this, domain adaptation techniques are employed. Let $D_S = \{(X_s^i, Y_s^i, Z_s^i)\}_{i=1}^{N_s}$ represent the source domain, encompassing N_s labeled training samples. Here, $X_s^i = \{X_t^i\}_{t=1}^{T_i} \in \mathbb{R}^{M_s \times T_i}$ denotes continuous operational condition settings, forming multivariate sequential data with a length of T_i and M_s setting variables. This signifies that for each instance i in the source domain, X_s^i is a matrix of size $M_s \times T_i$, capturing M_s different operational settings over a time sequence of T_i. $Y_s^i \in \mathbb{R}^{N_s \times T_i}$ captures sensory data in the feature space S. This means that for each instance i in the source domain, Y_s^i is a matrix of size $N_s \times T_i$, representing N_s sensory measurements over the same time sequence T_i. Z_s^i represents Remaining Useful Life (RUL) values with a length of T_i, indicating the remaining operational life for each time step t in the sequence. At each time step $t \in \{1, 2, ..., T_i\}$, the pair $[X_t^i, Y_t^i] \in \mathbb{R}^{M_s + N_s}$ and Z_t^i corresponds to the t-th CM measurement and RUL label, respectively.

Moving to the target domain, denoted as $D_T = \{(U_T^i, V_T^i)\}_{i=1}^{N_T}$, introduces a novel aspect in the form of $V_T^i \in \mathbb{W}_{\mathbb{T}}$ and $\mathbb{W}_{\mathbb{T}} \in \mathbb{R}^{R_T \times T_i}$. This implies that V_T^i represents a feature vector belonging to the space $\mathbb{W}_{\mathbb{T}}$, which encapsulates specific characteristics of the target environment. The matrix $\mathbb{W}_{\mathbb{T}}$ has dimensions $\mathbb{R}^{R_T \times T_i}$, signifying the feature space's dimensionality and the temporal sequence length for each instance in the target domain. Each column in $\mathbb{W}_{\mathbb{T}}$ represents a different feature, and each row represents a specific time point.

A critical assumption in this study involves asserting distinct marginal distributions for the source (D_S) and target (D_T) domains, explicitly expressed as $P(S) \neq P(\mathbb{W}_{\mathbb{T}})$. This assumption underscores the acknowledgment of inherent differences in data characteristics between the labeled training data and the unlabeled testing data, motivating the application of domain adaptation techniques to align these distributions effectively.

The primary objective of this research is to train a deep learning-based regression model F using the available datasets D_S and D_T to predict RULs accurately in the target domain. Formally, the model seeks to approximate $Z_T^i \approx F(X_T^i, Y_T^i)$. This study uniquely emphasizes engineered systems operating under dynamic time-varying conditions, aligning closely with practical scenarios in industrial settings. The challenges posed by dynamic operational conditions contribute practical relevance to the field of fault prognosis within the realm of Industry 4.0.

IV. Proposed Methodology

In the pursuit of predicting Remaining Useful Life (RUL), our research introduces a comprehensive methodology that leverages time-varying sensor data from k sensors $(x_1, x_2, ..., x_k)$ collected at discrete time intervals $(t_1, t_2, ..., t_n)$ as presented in Figure 1. The initial stage involves a Self-Attention mechanism, where weights $(\alpha_1 \cdot x_1, \alpha_2 \cdot x_2, ..., \alpha_k \cdot x_k)$ are assigned to each sensor based on their contributions to the overall system behavior. This allows the model to selectively focus on informative sensors, enhancing its capacity to discern crucial patterns within the data. Following the Self-Attention block, the processed data enters a Temporal Convolution Network (TCN), strategically designed to capture intricate temporal patterns. The TCN employs convolution operations over the temporal dimension, facilitating the extraction of salient features critical for accurate RUL predictions. Subsequently, the data undergoes a Squeeze & Excitation (SE) block, which amplifies informative features by recalibrating the significance of different channels within the data. This adaptive recalibration ensures that the model dynamically adjusts to the changing importance of different features, enhancing its ability to capture relevant information.

After the SE block, the output is flattened into a one-dimensional vector and passed through a Fully Connected (FC) layer. This layer refines the data, enabling the model to capture complex relationships and dependencies among different features. The sequential integration of these components contributes to a holistic representation of the system's temporal dynamics.

Ultimately, the processed data from the FC layer is utilized for predicting the Remaining Useful Life of the system. The model, having undergone training on a synthesis of temporal information, sensor interactions, and enhanced features, demonstrates proficiency in making precise predictions regarding the remaining operational lifespan of the system. This methodology, integrating advanced techniques in self-attention, temporal convolution, and feature recalibration, offers a robust framework for RUL prediction, capturing and extrapolating nuanced temporal patterns within the sensor data.
Sensor Selection Normalization Correlation Coefficient RUL Label

A. Data Preprocessing

1) Sensor Selection Using Box Plot Analysis: In the initial stages of data preprocessing for predictive modeling of

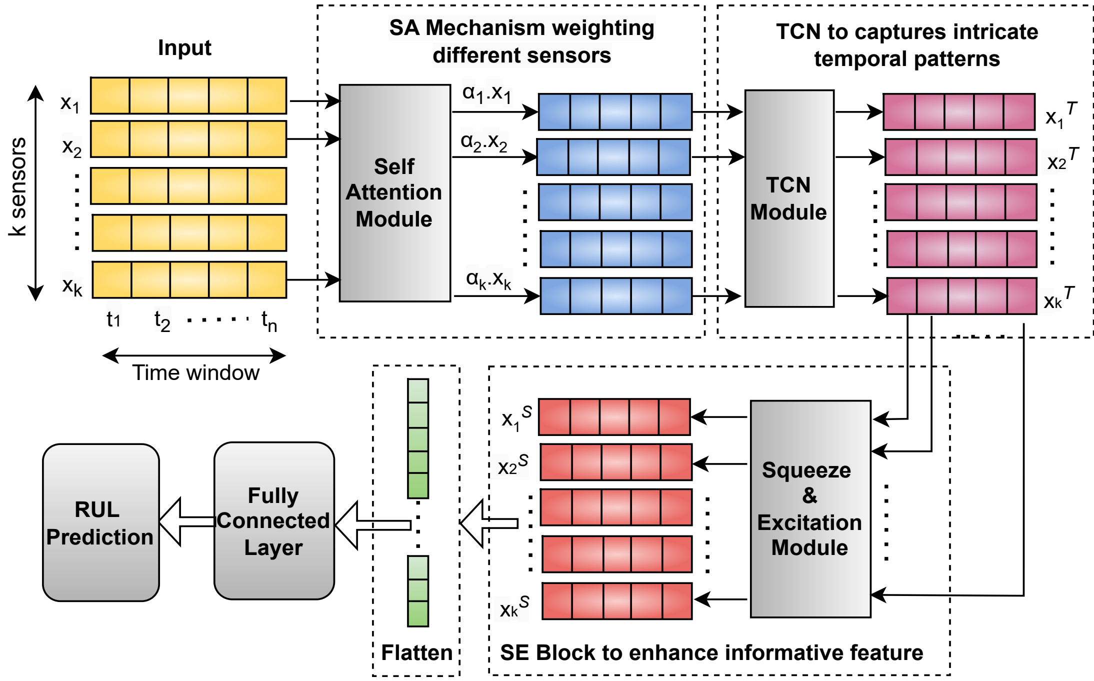

Fig. 1. Framework of the proposed method

Remaining Useful Life (RUL), a crucial step involved the meticulous selection of sensors based on insightful analysis using box plots. Box plots, also known as box-and-whisker plots, visually represent the distribution of data, providing a clear depiction of variability and trends within each sensor's measurements. This critical sensor selection process, outlined in Figure 2, aimed to enhance the relevance and effectiveness of the chosen sensors in predicting RUL while eliminating non-contributing sensors that may introduce noise to the model. As illustrated in Figure 2, sensors exhibiting clear data trends, namely sensor 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, and 21, were selected as input features for subsequent experiments. The decision to include these sensors was guided by the presence of distinct box plots showcasing variability in the data, indicating their capacity to capture meaningful information for RUL predictions. Conversely, sensors 1, 5, 6, 10, 16, 18, and 19 were excluded due to the absence of whiskers on either side or the absence of a box, indicative of a limited range of values. Sensors with no whiskers but retained, such as sensor 13, demonstrated the presence of outliers, signifying potential anomalies in the data that could contribute valuable insights to the predictive modeling process.

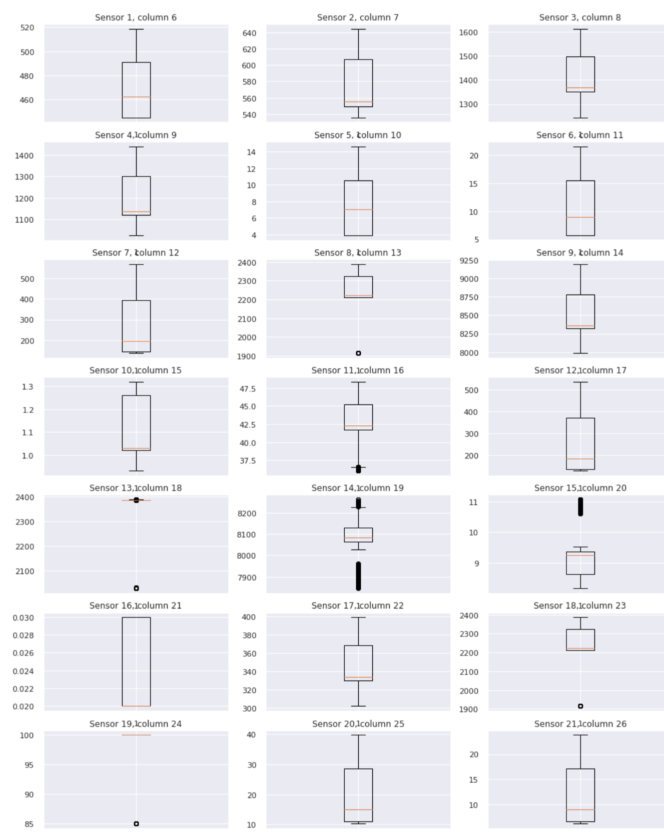

Fig. 2. Box plot analysis of sensor data.

2) Normalization of Multi-channel Sensor Data: In the analysis of multi-channel data collected from various sensors, addressing inherent variations in amplitude ranges is crucial. Each sensor corresponds to a unique channel, and the monitoring signals form a multi-channel dataset. To ensure comparability among diverse channel data, normalization is a critical preprocessing step. In this study, we adopt Min-Max normalization, also known as linear normalization, to scale the sensor data for each channel within the standardized range $[0, 1]$. The normalization process is defined by the Equation 1:

$$\hat{x}_{t,s} = \frac{x_{t,s} - \min(x_s)}{\max(x_s) - \min(x_s)} \tag{1}$$

Here, $\hat{x}_{t,s}$ denotes the normalized data for the s-th sensor at the t-th time step. The variables $x_{t,s}$, $\min(x_s)$, and $\max(x_s)$ represent the actual sensor data at time t, the minimum value of the sensor data, and the maximum value of the sensor data for the s-th sensor, respectively.
The Min-Max normalization transforms sensor data to a uniform scale, facilitating meaningful comparisons and analyses across different sensors. This normalization technique is crucial for ensuring that each sensor's contribution is appropriately weighted in subsequent analyses. It lays the foundation for robust and comprehensive feature extraction and modeling, contributing to the overall success of predictive modeling for Remaining Useful Life (RUL) predictions.

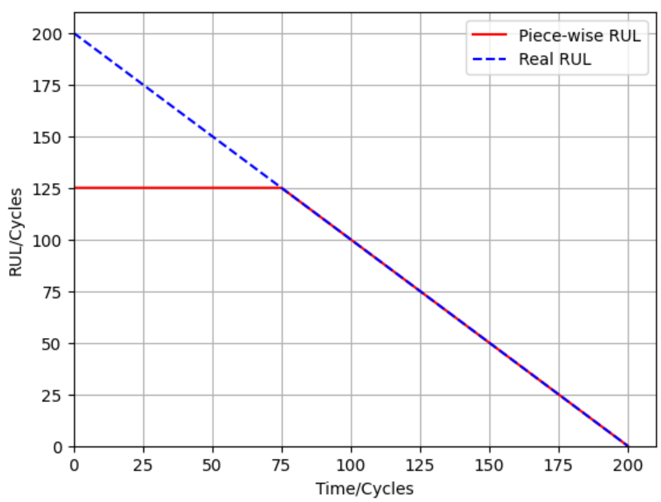

Fig. 3. RUL Label

3) RUL Label:

B. Multi Head Attention Block

The evolution of machine learning and deep learning techniques has been marked by continuous advancements in model architectures, enabling more nuanced processing of sequential data. One such pivotal development is the introduction of attention mechanisms, which revolutionized the field by allowing models to selectively focus on specific parts of input sequences. In this backdrop, the self-attention mechanism has emerged as a key element, offering unparalleled capabilities in capturing intricate relationships within sequences. The essence of self-attention lies in its ability to empower models to weigh the significance of different elements within a sequence dynamically as depicted in Figure 4. Unlike traditional models with fixed receptive fields, self-attention permits the model to assign varying degrees of importance to different parts of the input sequence, enabling it to discern complex dependencies. This mechanism has proven particularly effective in tasks involving long-range dependencies, as it mitigates the vanishing gradient problem encountered by traditional recurrent models. Mathematically, self-attention operates through the computation of attention scores, which determine the influence of one element over another within the sequence. The attention

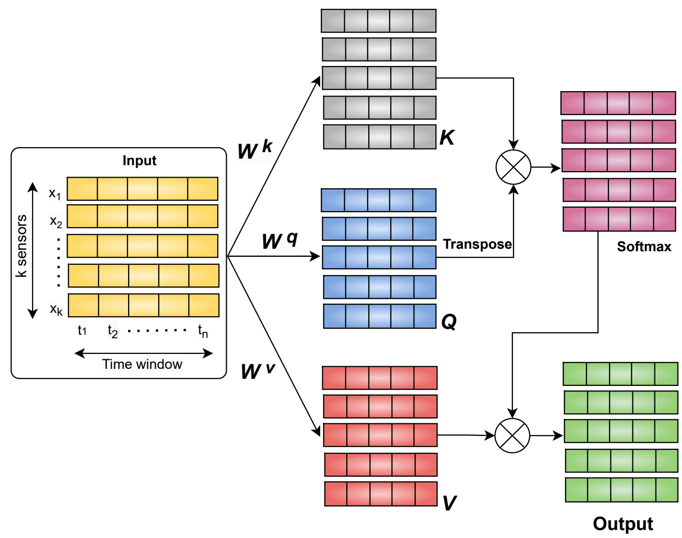

Fig. 4. Representation of Self Attention Module

scores are then utilized to compute a weighted sum of the input sequence, generating context-aware representations for each element. This process is encapsulated in the Equation 2:

$$\text{SA}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

Here, Q, K, and V represent the Queries, Keys, and Values, respectively, each derived from the input sequence. The softmax operation normalizes the attention scores, and the resulting weighted sum encapsulates the contextual information for each element. The self-attention mechanism has found widespread application in various domains, ranging from natural language processing tasks, such as machine translation, to sequential data analysis in predictive maintenance. Its adaptability to diverse contexts and its ability to capture both short-term and long-term dependencies make self-attention a cornerstone in modern deep-learning architectures. As we delve deeper into the complexities of predictive maintenance, we leverage the power of self-attention to capture nuanced patterns within sequential data. The ensuing sections will elaborate on the integration of self-attention within our proposed model, exploring its implications and contributions to the task of Remaining Useful Life (RUL) prediction.

The incorporation of SA block plays a crucial role in dynamically adapting to feature importance within degradation sequences. By utilizing convolutional layers for query, key, and value operations, the SA block allows the model to discern and prioritize significant temporal patterns. This adaptability proves essential for capturing nuanced variations in sensor data, addressing the intricate nature of degradation sequences inherent in the C-MAPSS dataset. The Softmax activation further refines the attention mechanism, emphasizing the importance of specific features in the context of RUL prediction. The integration of the SA block aligns with our objective of achieving superior predictive accuracy by enhancing the model's ability to discern subtle patterns and dependencies within the data.

In the pursuit of constructing a predictive maintenance model adept at addressing the independent and diverse nature of multidimensional equipment features in aircraft engine data, we introduce the Multi-Head Attention mechanism as a foundational component. Drawing inspiration from the transformer architecture, this attention mechanism is designed to accommodate the diverse and independent nature of multi-dimensional equipment features. At its core, the Multi-Head

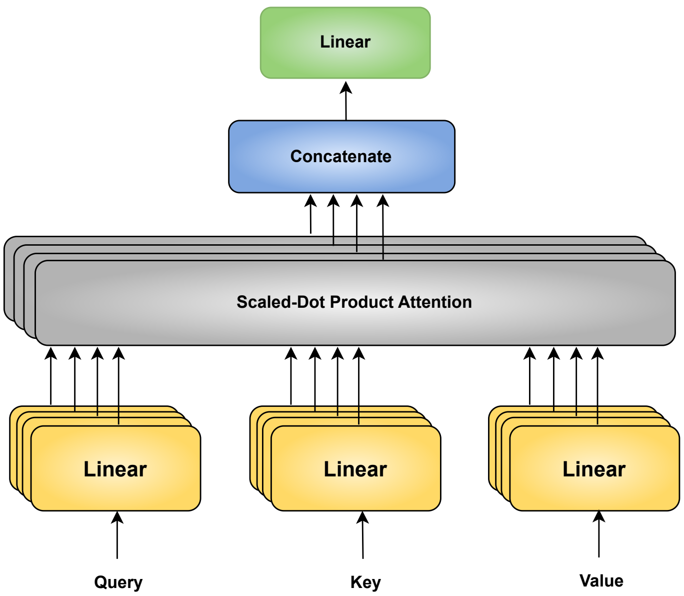

Fig. 5. Representation of Multi Head Attention Module

Attention mechanism enables our model to focus on different aspects of the input data simultaneously. This parallel processing capability is essential for capturing the rich and varied temporal dependencies present in the dynamic behavior of aircraft engines. Mathematically, the mechanism is expressed as follows:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_n)W_O, \quad (3)$$

where each head_i encapsulates the outcome of the attention function:

$$\text{head}_i = \text{Attention}(QW_{Qi}, KW_{Ki}, VW_{Vi}), \quad (4)$$

and W_O serves as the weight matrix facilitating concatenation. This mechanism ensures that the model not only attends to but comprehensively processes distinct features in parallel.

The attention function in each head is calculated as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (5)$$

where Q, K, and V denote the query, key, and value matrices, respectively, and d_k is the dimension of the key vectors.

The resulting output, a fusion of information from multiple attention heads, undergoes further refinement by a merge layer. This intricate process allows the model to capture and understand the interplay between different dimensions of the input data, providing a more holistic understanding of the underlying patterns.

Incorporating Multi-Head Attention into our methodology enhances the model's capacity to discern complex relationships, adapt to diverse features, and ultimately contributes to a more robust and accurate predictive maintenance system for estimating the Remaining Useful Life of aircraft engines.

C. *Temporal Convolution Network*

Temporal Convolutional Network (TCN), introduces a mechanism designed to capture temporal dependencies within sequential data. Unlike traditional recurrent architectures, TCN leverages dilated convolutions to expand the receptive field, allowing the model to effectively capture long-range dependencies without sacrificing computational efficiency as represented by Figure 6. The TCN architecture is defined by

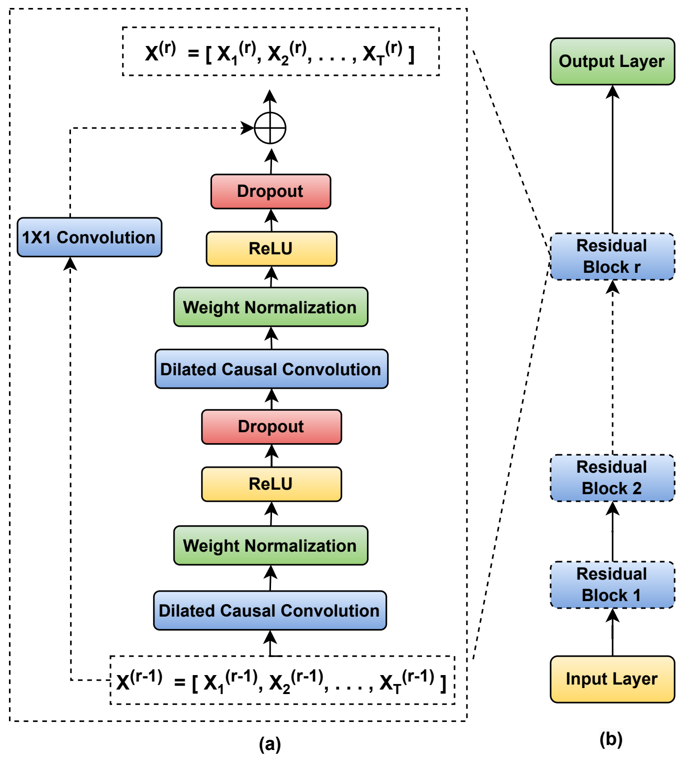

Fig. 6. Visualization of (a) Residual block. (b) Temporal convolutional network.

the convolutional layers, which operate sequentially on the input sequence. The integration of dilations enables TCN to incorporate information from distant past time steps while maintaining parallelizability. This is a significant advantage over recurrent models, where sequential dependencies often lead to increased computational complexity. The TCN mechanism involves the application of temporal convolutions to the input sequence, extracting features at different temporal scales. Mathematically, the operation of the TCN can be represented in Equation 6:

$$\text{TCN}(X; \Theta_{\text{TCN}}) = \text{ReLU}\left(\sum_{j=1}^{J} \Theta_{\text{TCN},j} * X\right) \quad (6)$$

Here, Θ_{TCN} represents the parameters associated with the TCN layer, and J denotes the number of dilated convolutional layers. The operation $*$ denotes the convolution operation. Moreover, TCNs integrate a generic residual module in place of convolutional layers to ensure stability in deeper and larger networks. The identity mapping in the residual module is expressed in Equation 7:

$$y = F(x, \{W_i\}) + x \tag{7}$$

Here, $F(x, \{W_i\})$ represents the residual mapping to be learned, enhancing the performance of very deep networks. The TCN's ability to capture temporal dependencies in parallel makes it a valuable component in our architecture for predictive maintenance. In the subsequent sections, we will explore the synergies between self-attention and TCN, elucidating their collective contributions to the task of Remaining Useful Life (RUL) prediction.

The Temporal Convolution Network (TCN) plays a pivotal role in our RUL estimation predictive maintenance model. Employed strategically after the Self-Attention (SA) block, TCN capitalizes on its efficiency in capturing long-range dependencies and overcoming the sequential limitations inherent in traditional Recurrent Neural Networks (RNNs). Comprising temporal blocks with two dilated convolutional layers, chomp, and dropout operations, the TCN architecture facilitates effective feature extraction. These TCN blocks are structured for parallelizable processing, enabling the model to efficiently capture temporal patterns within the degradation sequences of the C-MAPSS dataset. The decision to integrate TCN into our model is rooted in its superior ability to handle sequential data. The Chomp layer ensures causal convolution by removing the last elements from the sequence, allowing the convolution operation to consider only past information. Meanwhile, the dilated convolutions in the temporal blocks extend the network's receptive field without exponentially increasing parameters, facilitating the capture of intricate temporal dependencies. This design choice not only enhances predictive accuracy but also addresses the limitations associated with sequential computation in conventional RNNs. The adaptability and efficiency of TCN in capturing temporal dependencies solidify its significance within our model's architecture, contributing substantially to the overall effectiveness in predicting RUL for the challenging C-MAPSS dataset.

D. Squeeze & Excitation Block

The Squeeze & Excitation (SE) Block constitutes a pivotal element in our model, employed after the TCN output, to enhance feature recalibration and global contextual understanding. This strategic integration aims to address the limitations of conventional Convolutional Neural Networks (CNNs) by selectively emphasizing informative features.

In the domain of feature transformation, a Squeeze-and-Excitation (SE) block serves as a foundational computational unit. This block is built upon a transformation F_{tr} that maps an input $X \in \mathbb{R}^{n_0 \times c_0 \times t_0}$ to feature maps $U \in \mathbb{R}^{n \times c \times t}$. To maintain simplicity in notation, we assume F_{tr} as a convolutional operator, and $W = [w_1, w_2, ..., w_t]$ symbolizes the learned set of filter kernels, where w_t denotes the parameters of the t-th filter. The resulting outputs are denoted as $U = [u_1, u_2, ..., u_t]$, given by Equation 8

$$u_t = w_t * X = \sum_{s=1}^{t_0} w_s^t * x_s \tag{8}$$

Here, the $*$ signifies convolution, $w_t = [w_t^1, w_t^2, ..., w_{t_0}^{T'}]$, $X = [x^1, x^2, ..., x^{T'}]$, and $u_t \in \mathbb{R}^{n \times c}$. The term w_t^s represents a 2D spatial kernel capturing a single channel of w_t that operates on the corresponding channel of X. For conciseness, bias terms are excluded in this representation. Although channel dependencies are implicitly embedded in w_c through the summation across all channels, they intertwine with local spatial correlations captured by the filters. To enhance the learning of convolutional features, we advocate for explicitly modeling channel interdependencies, enabling the network to heighten sensitivity to informative features, thus making them more exploitable by subsequent transformations. As a

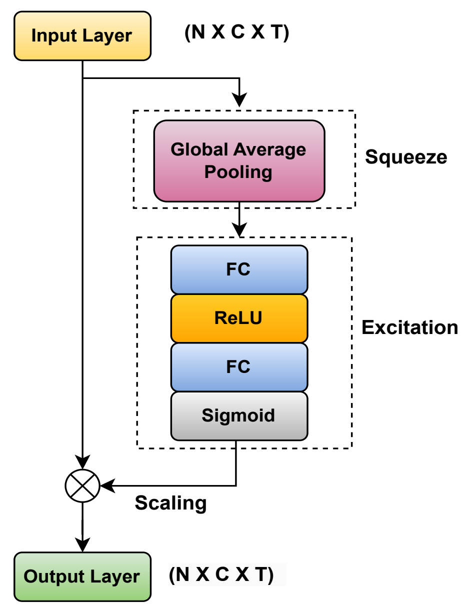

Fig. 7. Illustration of Squeeze & Excitation Module

consequential step, we introduce the Squeeze-and-Excitation (SE) mechanism as visualized in Figure 7, to provide the network with global information access and recalibrate filter responses through two steps: squeeze and excitation. The objective is to dynamically adjust attention across each channel before propagating them into the subsequent transformation. Mathematically, the SE mechanism is articulated in Equation 9:

$$\text{SEBlock}(X; \Theta_{\text{SE}}) = \sigma\left(\Theta_{\text{SE}_2} \delta(\Theta_{\text{SE}_1} \text{AvgPool}(X))\right) \odot X \tag{9}$$

Here, Θ_{SE_1} and Θ_{SE_2} denote learnable weights, AvgPool denotes global average pooling, δ represents the Rectified Linear Unit (ReLU) activation function, σ is the sigmoid activation function, and $\odot$ signifies element-wise multiplication. The introduction of the Squeeze-and-Excitation mechanism enriches the network's capacity to adaptively attend to essential features, contributing to improved feature transformation and, consequently, enhanced fault prognosis capabilities. The explicit modeling of channel interdependencies aligns with the network's ability to capture both local and global contextual information. This integration aims to elevate the model's sensitivity to informative features critical for accurate Remaining Useful Life (RUL) predictions.

The SE Block operates through a two-step mechanism: squeezing and exciting. The squeezing phase involves adaptive average pooling to compress the spatial dimensions, condensing the input features. Subsequently, the excitation phase employs linear transformations and activation functions, specifically ReLU and Sigmoid, to recalibrate and highlight essential features based on their significance. The decision to incorporate the SE Block is motivated by its efficacy in enhancing global contextual understanding. By emphasizing critical features and suppressing less informative ones, the SE Block enables the model to focus on key information, thereby refining predictions. The choice of SE Block aligns with our objective of optimizing the model's performance in predicting Remaining Useful Life (RUL) for the C-MAPSS dataset. The SE Block's architecture, with its adaptive attention mechanism, contributes to the overall robustness and accuracy of our predictive maintenance model. This adaptive attention mechanism, combined with ReLU activation and linear transformations, serves as a sophisticated mechanism for feature recalibration, addressing nuances in the dataset and further elevating the model's predictive capabilities. The utilization of the SE Block in our model signifies a deliberate choice to augment global contextual understanding, effectively addressing information bottlenecks and enhancing the model's capacity to discern critical patterns within the C-MAPSS dataset.

E. Model Optimization

The optimization process for our predictive maintenance model is a pivotal step aimed at enhancing its predictive performance for Remaining Useful Life (RUL). The model architecture incorporates a Multi Head Attention (MHA) layer, a Temporal Convolutional Network (TCN), and a Squeeze-and-Excitation (SE) block, strategically designed to capture dynamic dependencies within the dataset. The optimization objective is formulated by minimizing the loss function defined as:

$$\min_{\Theta} \mathcal{L}(\Theta) = \frac{1}{M} \sum_{i=1}^{M} \sqrt{\text{MSE}(y_i, \hat{y}_i; \Theta)} \tag{10}$$

Here, Θ represents the model parameters, including MHA (Θ_{MHA}), TCN (Θ_{TCN}), and SE (Θ_{SE}). M denotes the total number of samples, y_i is the true RUL for the i-th sample, $\hat{y}_i$ is the predicted RUL for the i-th sample, and $\text{MSE}(\cdot)$ is the Mean Squared Error loss function. The optimization process seeks to find the optimal set of parameters Θ that minimizes the Root Mean Squared Error (RMSE) loss across the entire dataset. During the training process, hyperparameters are carefully tuned to optimize the model's performance. This involves adjusting parameters such as learning rates, batch sizes, and the number of training epochs. The architecture of the predictive maintenance model is expressed in Equation 11:

$$\hat{Y} = \text{SE}\left(\text{TCN}\left(\text{MHA}(X; \Theta_{\text{MHA}}); \Theta_{\text{TCN}}\right); \Theta_{\text{SE}}\right) \tag{11}$$

Here, $\hat{Y}$ denotes the predicted RUL for a given input X, X represents the input data at a specific time step, and Θ_{MHA}, Θ_{TCN}, and Θ_{SE} are parameters associated with the SA layer, TCN layer, and SE block, respectively. This holistic model architecture, coupled with the optimization process defined by Equation (10), is meticulously designed to elevate the accuracy and effectiveness of our predictive maintenance model in estimating the Remaining Useful Life. The optimization aims to adaptively adjust the model parameters to minimize prediction errors and improve its overall predictive capability.

F. Prognosis Procedure

In our novel model for predicting Remaining Useful Life (RUL), we present a structured methodology encompassing key stages such as data input, preprocessing, model training, testing, and evaluation. The schematic diagram in Figure 8 visually illustrates the sequential flow, offering clarity on the essential steps integral to our approach. The methodology

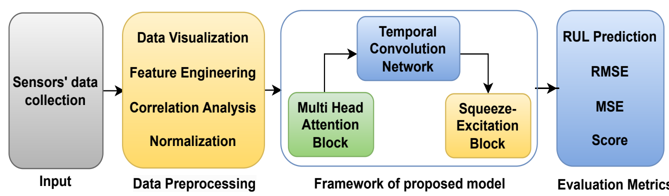

Fig. 8. Overview of the flow of the proposed methodology

begins with data input, where sensor data from the dataset is introduced into the system. A robust data preprocessing step follows to ensure the quality and relevance of the input data. During data preprocessing, we strategically select sensors based on variance, employ exploratory data analysis with box plots and heatmaps for feature refinement and selection, and critically normalize data using a min-max scaler. These steps collectively enhance the dataset's quality and relevance, laying the groundwork for the development of an accurate RUL prediction model. After data preprocessing, the model training phase utilizes the proposed architecture, which includes a Self-Attention (SA) block, Temporal Convolution Network (TCN), and Squeeze and Excitation (SE) block. The testing phase assesses the model's performance on unseen data, and the evaluation employs metrics such as Root Mean Square Error (RMSE) and score. These metrics serve as quantitative measures of the model's accuracy and effectiveness in predicting RUL. This architecture is meticulously crafted to address the limitations of existing models in predicting RUL, providing a comprehensive and advanced solution to the RUL prediction problem. The proposed methodology is designed to enhance predictive accuracy and overcome challenges associated with existing approaches.

V. Performance Evaluation

A. Dataset Description

The C-MAPSS Turbofan Engine Simulation Dataset [Add Reference], released by NASA, has been widely used for evaluation of fault prognostic models. The dataset consist of four data sub-sets FD001, FD002, FD003 and FD004 which are single fault single mode, single fault multi mode, multimode single fault and multi fault multimode. This dataset, originating from 218 turbofan jet engines, comprises run-to-failure simulated data collected from 21 attached sensors. These sensors generate multivariate time-series data organized in X-by-26 matrices, where X represents the number of samples in the subset. Each sample is a 26-dimensional vector encompassing engine number, cycle time, three operating conditions, and 21 run-to-failure sensor signals. The objective of the dataset is to predict the RUL of engines, with actual RUL labels provided in the test data for validation. The dataset's versatility is showcased through its ability to simulate various fault and deterioration degrees in turbofan engines, making it a valuable resource for predictive maintenance studies.

B. Dataset Preprocessing & Experimental Setup

In C-MAPSS Dataset, the sensors with minimal variance are excluded and focusing on remaining sensors input that contribute significantly. For uniform contribution from all features, both sensor data and operating condition data are normalized using min-max scaler before feeding to proposed model for RUL prediction. Further, the tuned hyperparameters in proposed model are detailed in Table I. The tuning of hyperparameters aims to enhance the model's ability to discern subtle patterns indicative of engine degradation, facilitating precise predictions of RUL across diverse operational scenarios. These hyperparameters serve as critical elements in shaping the model's capability to provide accurate and reliable predictions.

TABLE I
Hyperparameters for the proposed model for C-MAPSS FD001, FD002, FD003, and FD004 dataset

Parameter	Values
batch_size	256, 512, 512, 512
dropout	0.2, 0.2, 0.2, 0.2
epochs	120, 200, 200, 200
gamma_scheduler	0.02, 0.1, 0.02, 0.1
kernel_size	3, 3, 3, 3
learning_rate	0.01, 0.01, 0.01, 0.01
milestones_scheduler	40, [100, 180], 40, [100, 180]
num_channels	[16, 16, 16], [16, 16], [16, 16], [16,16]
num_samples	100, 259, 100, 248
output_size	1, 1, 1, 1
reduction_ratio (r)	2, 2, 2, 2
sensor_num	14, 17, 14, 17
tcn_inputs_size	14, 17, 14, 17
window_size	31, 40, 45, 50

To evaluate the performance of proposed predictive maintenance, two metrics namely Root Mean Square Error (RMSE) and score metric are considered. RMSE is widely adopted symmetrical scoring function that measures the average magnitude of prediction errors between the actual and predicted RUL values. Mathematically, RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N}} \tag{12}$$

Here, e_i represents the prediction error for each sample, and N is the total number of samples. RMSE assigns equal weights or penalties to both early and late predictions and fails to distinguish under-estimation and over-estimation. To address this issue, score metric is considered for asymmetric evaluation function that assigns larger penalty for over-estimation. Mathematically, score metric can be expressed as

$$\text{score} = \sum_{j=1}^{N} s_j \tag{13}$$

where s_j is calculated based on the prediction error (e) and a predefined function h_j

$$s_j = \begin{cases} e^{-h_j/13} - 1, & \text{if } h_j < 0 \\ e^{-h_j/10} - 1, & \text{if } h_j \geq 0 \end{cases} \tag{14}$$

The parameter h_j represents the difference between actual and predicted RUL. This scoring function accentuates the impact of delayed predictions, reflecting the real-world consequences of such deviations.

C. Experimental Result and Analysis

In the pursuit of refining our predictive maintenance model, we conducted a meticulous analysis by deconstructing the proposed architecture into three distinctive configurations: MHA+TCN, TCN+SE, and MHA+TCN+SE. Each configuration represents a deliberate alteration in the model's architecture to discern the individual impact of the Multi-Head Attention (MHA), Temporal Convolution Network (TCN), and Squeeze-and-Excitation (SE) components.

1) Significance of MHA+TCN in Predicting RUL: The combination of Multi-Head Attention (MHA) and Temporal Convolution Network (TCN) stands out as a significant architectural arrangement. This integration leverages the strengths of attention mechanisms and temporal convolution, providing a comprehensive method to capture detailed temporal dependencies and intricate patterns in machinery degradation. The Multi-Head Attention mechanism excels in recognizing nuanced temporal relationships, enabling the model to gain detailed insights into the Remaining Useful Life (RUL). At the same time, the Temporal Convolution Network is adept at capturing complex temporal patterns, contributing to a holistic understanding of machinery degradation over time. The collaborative nature of MHA and TCN ensures a synergistic effect, with MHA focusing on specific temporal dependencies and TCN capturing broader temporal patterns. This collaboration establishes an architecture well-suited to address the challenges of predictive maintenance tasks. The hyperparameter

window_size varied for FD001, FD002, FD003, and FD004, assuming values of 14, 17, 14, and 14, respectively.
Table II presents the quantitative metrics, including RMSE, Score, and Accuracy, for each sub-dataset (FD001, FD002, FD003, FD004) using the MHA+TCN configuration. The Score represents the performance of the last sample in the test set. The MHA+TCN configuration exhibits optimal RMSE values and accuracy across all sub-datasets compared to the other general predictive models.

TABLE II
QUANTITATIVE ANALYSIS OF MHA+TCN CONFIGURATION: RMSE, SCORE, AND ACCURACY FOR ENGINE SUB-DATASETS

	FD001	FD002	FD003	FD004
RMSE	14.13	17.35	12.8	15.73
Score	6.98×10^2	1.6×10^3	6.25×10^2	4.1×10^3
Accuracy	0.67	0.54	0.75	0.68

In Figure 9, column (a) illustrates the loss with epoch, where the x-axis denotes the number of epochs, and the y-axis represents the RMSE loss. Fluctuations in the graph suggest dynamic convergence during the training process. The model exhibits acceptable performance, albeit with some variability. Moving to column (b), individual sample results for the MHA+TCN configuration are presented. The x-axis denotes sample numbers, while the y-axis displays RUL values. Blue lines indicate predicted RUL values, and red lines represent true RUL values. Divergence between predicted and true values suggests a reduction in accuracy. In column (c), specific sample results for sample numbers (100 for FD001 and FD003, 248 for FD002 and FD004) using the MHA+TCN configuration are showcased. The x-axis represents time cycles, and the y-axis denotes RUL values. Although predicted and true values remain aligned, a slight deviation implies a modest decrease in accuracy.

2) Significance of TCN+SE in Predicting RUL: The integration of Temporal Convolution Network (TCN) and Squeeze-and-Excitation (SE) configuration represents a noteworthy exploration in predictive maintenance models. TCN excels in capturing complex temporal patterns, providing a robust approach to understanding machinery degradation over time. The addition of the SE block introduces an extra layer of sophistication, emphasizing critical features and contributing to enhanced predictive accuracy. TCN, in collaboration with SE, offers a balance between capturing intricate temporal patterns and emphasizing critical features. The Temporal Convolution Network captures broad temporal patterns, while the Squeeze-and-Excitation block adds a layer of sophistication by focusing on critical features. This combined approach creates an architecture designed to address the nuanced aspects of machinery degradation in predictive maintenance tasks. The hyperparameter window_size varied for FD001, FD002, FD003, and FD004, taking on values of 14, 17, 14, and 17, respectively.
Table III presents the quantitative metrics, including RMSE, Score, and Accuracy, for each sub-dataset (FD001, FD002, FD003, FD004) using the TCN+SE configuration. The Score represents the performance of the last sample in the test set. The TCN+SE configuration demonstrates a noticeable

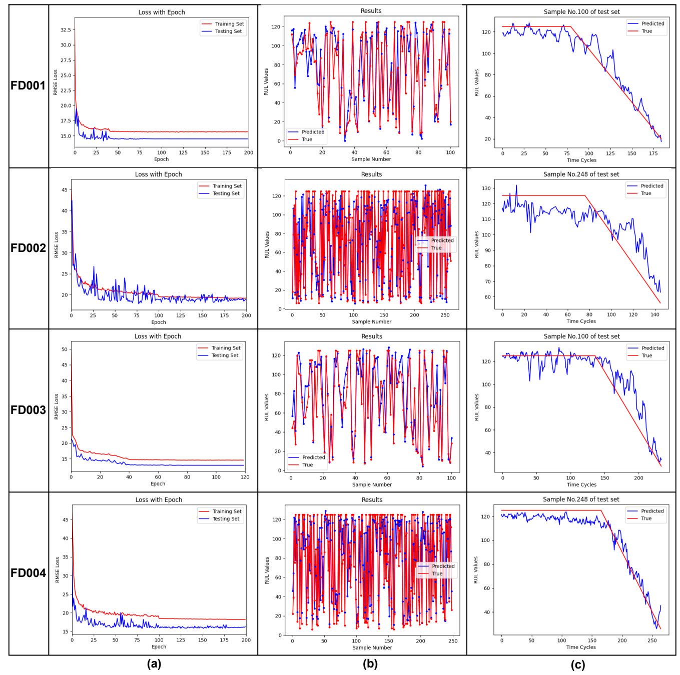

Fig. 9. Visualizing MHA+TCN Model Results: Loss with Epoch, Time-Cycle Specifics, and Individual Samples

decrease in accuracy across all sub-datasets, emphasizing the critical role of Multi-Head Attention in enhancing predictive capabilities.

TABLE III
QUANTITATIVE ANALYSIS OF TCN+SE CONFIGURATION: RMSE, SCORE, AND ACCURACY FOR ENGINE SUB-DATASETS

	FD001	FD002	FD003	FD004
RMSE	14.35	20.71	13.9	18.16
Score	9.11×10^2	3.09×10^3	8.47×10^2	5.7×10^3
Accuracy	0.65	0.43	0.69	0.50

Figure 10 illustrates the results of the TCN+SE configuration, highlighting the impact of excluding Multi-Head Attention. In column (a), the loss with epoch is depicted, where the x-axis represents the number of epochs, and the y-axis signifies the RMSE loss. The observed fluctuations in the graph suggest a less stable convergence compared to the MHA+TCN configuration. The absence of MHA introduces additional variability in the training process, resulting in higher RMSE values.
Column (b) of Figure 10 showcases individual samples' results for the TCN+SE configuration. The x-axis represents sample numbers, while the y-axis displays RUL values. Blue lines denote predicted RUL values and red lines represent true RUL values. The divergence between predicted and true values indicates a reduction in accuracy compared to the configurations involving Multi-Head Attention.
Column (c) of Figure 10 presents the results for specific sample numbers (100 for FD001 and FD003, 248 for FD002 and FD004) using the TCN+SE configuration. The x-axis represents time cycles, and the y-axis denotes RUL values. While the predicted and true values are still aligned, the slight deviation suggests a moderate decrease in accuracy. The

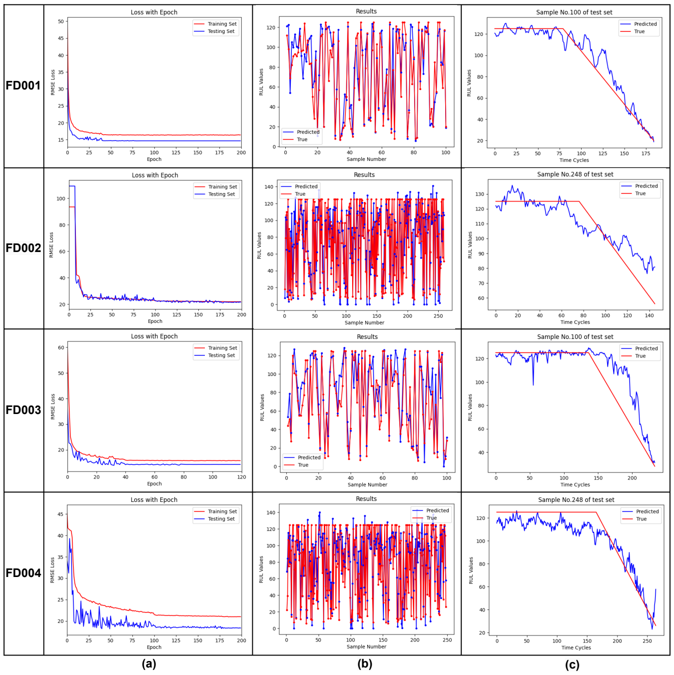

Fig. 10. Visualizing TCN+SE Model Results: Loss with Epoch, Time-Cycle Specifics, and Individual Samples

TCN+SE configuration, lacking the MHA module, demonstrates a commendable performance but falls short of the heightened accuracy achieved by configurations involving Multi-Head Attention.

3) Significance of MHA+TCN+SE in predicting RUL: Within the comprehensive MHA+TCN+SE configuration, our model demonstrated a remarkable fusion of architectural components, leading to superior predictive capabilities. The Multi-Head Attention (MHA) mechanism exhibited adeptness in discerning intricate temporal dependencies, enabling nuanced insights into the Remaining Useful Life (RUL). Leveraging these insights, the Temporal Convolution Network (TCN) adeptly captured complex temporal patterns. The Squeeze-and-Excitation (SE) block added an extra layer of sophistication, accentuating critical features and contributing to heightened predictive accuracy. The resultant model not only achieved the lowest Root Mean Square Error (RMSE) and score but also secured the highest accuracy, firmly establishing MHA+TCN+SE as the optimal configuration for predictive maintenance tasks. The selected hyperparameter, window_size, varied across experiments for FD001, FD002, FD003, and FD004, taking on values of 31, 40, 45, and 50, respectively. Table IV presents the quantitative metrics, including RMSE, Score, and Accuracy, for each sub-dataset (FD001, FD002, FD003, FD004) using the MHA+TCN+SE configuration. The Score represents the performance of the last sample in the test set. The MHA+TCN+SE configuration demonstrates consistent and high-performance metrics across all sub-datasets, with particularly noteworthy results for FD003, where the RMSE is minimal, and accuracy reaches 81%.

Figure 11 shows the result of the MHA+TCN+SE configuration, where the row-wise result is shown for FD001, FD002, FD003, and FD004. The column (a) of Figure 11 represents the loss with epoch for the MHA+TCN+SE configuration. The x-axis denotes the number of epochs, and the y-axis illustrates the RMSE loss. The smooth curve observed in the graph is a testament to the effectiveness of the MHA+TCN+SE configuration. Notably, in FD003, the model achieves optimal performance with training converging at 120 epochs, compared to other datasets where training extends up to 200 epochs. This suggests that in FD003, the model reaches a point of optimal accuracy, and further training might lead to overfitting. The convergence patterns in the loss graph indicate the efficiency of the MHA+TCN+SE configuration. The optimal training epochs vary across datasets, emphasizing the adaptive nature of the model to dataset-specific characteristics. The column (b) of Figure 11 depicts the individual samples' results for the MHA+TCN+SE configuration. The x-axis represents sample numbers, and the y-axis displays RUL values. Blue lines signify predicted RUL values, while red lines represent true RUL values. The close alignment of blue (predicted) and red (true) lines in the individual samples plot indicates accurate predictions across different subsets. The model effectively captures the nuances in RUL variations for specific samples. The column (c) of Figure 11 illustrates the results for specific

TABLE IV
QUANTITATIVE ANALYSIS OF MHA+TCN+SE CONFIGURATION: RMSE, SCORE, AND ACCURACY FOR ENGINE SUB-DATASETS

	FD001	FD002	FD003	FD004
RMSE	10.62	15.21	9.28	13.18
Score	1.90×10^2	1.09×10^3	2.35×10^2	1.12×10^3
Accuracy	0.75	0.67	0.83	0.72

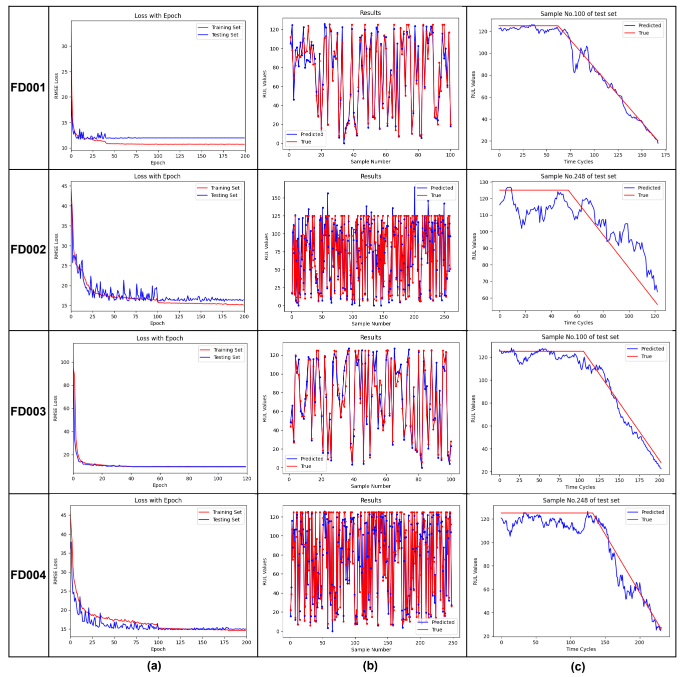

Fig. 11. Visualizing MHA+TCN+SE Model Results: Loss with Epoch, Time-Cycle Specifics, and Individual Samples

sample numbers (100 for FD001 and FD003, 248 for FD002 and FD004) using the MHA+TCN+SE configuration. The x-axis represents time cycles, and the y-axis denotes RUL values. The overlapping lines of predicted and true values signify accurate RUL estimation. The MHA+TCN+SE model excels in capturing the dynamic degradation patterns over time, enhancing the overall predictive accuracy.

D. Comparative analysis with other predictive models

To underline the robustness and superiority of our proposed model, we conducted a comprehensive comparative analysis, assessing its performance against contemporary prognostic methodologies designed for the C-MAPSS dataset. The detailed results presented in Table V underscore the substantial advancements achieved by our model across sub-datasets FD001, FD002, FD003, and FD004 when compared to established methods such as CNN, LSTM, BiLSTM, DAG, CNN+LSTM, Multi-head CNN+LSTM, CNN+LSTM+BiLSTM, AGCNN, LSTM+FCLCNN, Hybrid model, BLS + TCN, Bi-LSTM based Attention method, and LSTM without automatic piece-wise linear RUL function. The superior performance of our model is attributed to its innovative architectural design, seamlessly integrating the Multi-Head Attention (MHA) block, Temporal Convolution Network (TCN), and Squeeze-and-Excitation (SE) block. This synergistic combination empowers the model with dynamic feature adaptation, efficient capture of long-range dependencies, and enhanced global contextual understanding, effectively addressing the inherent complexities of prognostic tasks.

In the comparative analysis, traditional methods like CNN, LSTM, and BiLSTM exhibited varying degrees of effectiveness but consistently fell short when compared to our proposed model. CNN-based approaches displayed elevated RMSE values, indicating limitations in capturing intricate temporal dependencies. LSTM and BiLSTM models, while delivering competitive performance, lacked the ability to adapt dynamically to varying feature importance. DAG, though showcasing competitive RMSE, lagged in terms of score metrics.

Hybrid approaches such as CNN+LSTM and Multi-head CNN+LSTM demonstrated competitive yet inferior performance compared to our model. The AGCNN method presented notable metrics but was outperformed across all sub-datasets by our proposed model. The LSTM+FCLCNN approach presented incomplete data, yet the available metrics suggested suboptimal performance. The Hybrid model yielded mixed results, emphasizing the need for a more tailored approach.

Additionally, methodologies like BLS + TCN and Bi-LSTM based Attention showed competitive yet inferior metrics when compared to our proposed model. The LSTM without an automatic piece-wise linear RUL function exhibited suboptimal performance, suggesting that the LSTM network, in the absence of an automatic mechanism for adapting the RUL estimation process, may encounter difficulties in accurately modeling degradation patterns and predicting RUL values.

Our proposed model consistently outperformed all comparative methods across sub-datasets, showcasing lower RMSE values and higher score metrics. This superior performance establishes the efficacy of our model in addressing the challenges associated with prognostic tasks, particularly in the context of the C-MAPSS dataset. The chosen hyperparameters, including the window_size, were carefully tuned to ensure optimal performance across FD001, FD002, FD003, and FD004 sub-datasets. The comprehensive evaluation presented in Table V substantiates the effectiveness of our proposed model and its suitability for prognostic tasks in diverse operational scenarios.

VI. Conclusion

In conclusion, this research has endeavored to advance the field of Remaining Useful Life (RUL) prediction in the context of industrial machinery, specifically focusing on the C-MAPSS turbofan engine degradation dataset. The study incorporated meticulous data preprocessing techniques, including sensor selection through correlation analysis and data refinement using a moving median filter. The integration of a novel predictive maintenance model, SAFETNet (Self-Attention + Temporal Convolutional Network + Squeeze & Excitation), demonstrated superior performance compared to established models, particularly LSTM, BiLSTM, and CNN, as evidenced by minimal Root Mean Square Error (RMSE) and optimal score metrics.

Despite the promising results, this study has certain limitations. The performance of SAFETNet is contingent on the quality and relevance of the input data. Addressing challenges related to multi-sensor data fusion and environmental noise remains an ongoing endeavor. Additionally, the proposed model's adaptability and its generalizability to other industrial contexts require further investigation.

Future research should delve into enhancing the robustness of SAFETNet by developing advanced data preprocessing techniques for diverse multi-sensor data. Additionally, investigating adaptive algorithms to mitigate environmental noise and exploring the integration of SAFETNet with other fault prediction algorithms could contribute to more accurate and adaptable RUL predictions. Furthermore, extending the applicability of SAFETNet to different industrial domains and evaluating its performance under varied operating conditions will be crucial for its widespread adoption. This study paves the way for future research endeavors aimed at refining predictive maintenance models for improved industrial machinery prognosis.

Acknowledgment

References

[1] Y.-C. Liu, Y.-J. Chang, S.-L. Liu, and S.-P. Chen, "Data-driven prognostics of remaining useful life for milling machine cutting tools," in *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE, 2019, pp. 1–5.

[2] C. Liu, D. Tang, H. Zhu, and Q. Nie, "A novel predictive maintenance method based on deep adversarial learning in the intelligent manufacturing system," *IEEE access*, vol. 9, pp. 49 557–49 575, 2021.

[3] J. J. M. Jimenez, S. Schwartz, R. Vingerhoeds, B. Grabot, and M. Salaün, "Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics," *Journal of manufacturing systems*, vol. 56, pp. 539–557, 2020.

TABLE V
COMPARATIVE PERFORMANCE RESULTS FOR PROGNOSTIC METHODS ON THE C-MAPSS DATASET

Methods	FD001		FD002		FD003		FD004	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
CNN[41]	18.44	1.29×10^3	30.29	1.36×10^4	19.81	1.60×10^3	29.15	7.89×10^3
LSTM [38]	16.14	3.38×10^2	24.49	4.45×10^3	16.18	8.52×10^2	28.17	5.55×10^3
BiLSTM [62]	13.65	2.95×10^2	23.18	4.13×10^3	13.74	3.17×10^2	24.86	5.43×10^3
DAG [63]	11.96	2.29×10^2	20.34	2.73×10^3	12.46	5.35×10^2	22.43	3.37×10^3
CNN+LSTM [64]	16.16	3.03×10^2	20.44	3.44×10^3	17.12	1.42×10^3	23.25	4.63×10^3
Multi-head CNN+LSTM [45]	12.19	2.59×10^2	19.93	4.35×10^3	12.85	3.43×10^2	22.89	4.34×10^3
CNN+LSTM+ BiLSTM [44]	10.41	-	-	-	-	-	-	-
AGCNN [65]	12.42	2.25×10^2	19.43	1.49×10^3	13.39	2.27×10^2	21.50	3.39×10^3
LSTM+ FCLCNN [66]	11.17	2.04×10^2	-	-	9.99	2.34×10^2	-	-
Hybrid model [67]	15.68	-	22.26	-	16.89	-	22.32	-
BLS + TCN [68]	12.08	2.43×10^2	16.87	1.60×10^3	11.43	2.44×10^2	18.12	2.09×10^3
Bi-LSTM based Attention method [69]	13.78	2.55×10^2	15.94	1.28×10^3	14.36	4.38×10^2	16.96	1.65×10^3
LSTM without automatic piece-wise linear RUL function	13.5	2.38×10^2	23.37	2.6×10^3	13.54	4.11×10^2	23.36	3.97×10^3
DATCN	-	-	15.95	1.158×10^3	-	-	18.64	1.806×10^3
Proposed	10.62	1.90×10^2	15.21	1.09×10^3	9.28	2.35×10^2	13.18	1.12×10^3

[4] S. Behera, A. Choubey, C. S. Kanani, Y. S. Patel, R. Misra, and A. Sillitti, "Ensemble trees learning based improved predictive maintenance using iiot for turbofan engines," in *Proceedings of the 34th ACM/SIGAPP symposium on applied computing*, 2019, pp. 842–850.

[5] K. T. Nguyen and K. Medjaher, "A new dynamic predictive maintenance framework using deep learning for failure prognostics," *Reliability Engineering & System Safety*, vol. 188, pp. 251–262, 2019.

[6] C. Zheng, W. Liu, B. Chen, D. Gao, Y. Cheng, Y. Yang, X. Zhang, S. Li, Z. Huang, and J. Peng, "A data-driven approach for remaining useful life prediction of aircraft engines," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 184–189.

[7] V. Mathew, T. Toby, V. Singh, B. M. Rao, and M. G. Kumar, "Prediction of remaining useful lifetime (rul) of turbofan engine using machine learning," in *2017 IEEE international conference on circuits and systems (ICCS)*. IEEE, 2017, pp. 306–311.

[8] N. Bolander, H. Qiu, N. Eklund, E. Hindle, and T. Rosenfeld, "Physics-based remaining useful life prediction for aircraft engine bearing prognosis," in *Annual Conference of the PHM Society*, vol. 1, no. 1, 2009.

[9] C. Xiongzi, Y. Jinsong, T. Diyin, and W. Yingxun, "Remaining useful life prognostic estimation for aircraft subsystems or components: A review," in *Ieee 2011 10th international conference on electronic measurement & instruments*, vol. 2. IEEE, 2011, pp. 94–98.

[10] P. Adhikari, H. G. Rao, and M. Buderath, "Machine learning based data driven diagnostics & prognostics framework for aircraft predictive maintenance," in *Proceedings of the 10th International Symposium on NDT in Aerospace, Dresden, Germany*, 2018, pp. 24–26.

[11] L. Liu, Q. Guo, D. Liu, and Y. Peng, "Data-driven remaining useful life prediction considering sensor anomaly detection and data recovery," *IEEE access*, vol. 7, pp. 58 336–58 345, 2019.

[12] M. Aykol, C. B. Gopal, A. Anapolsky, P. K. Herring, B. van Vlijmen, M. D. Berliner, M. Z. Bazant, R. D. Braatz, W. C. Chueh, and B. D. Storey, "Perspective—combining physics and machine learning to predict battery lifetime," *Journal of The Electrochemical Society*, vol. 168, no. 3, p. 030525, 2021.

[13] W. Zhang, D. Yang, and H. Wang, "Data-driven methods for predictive maintenance of industrial equipment: A survey," *IEEE systems journal*, vol. 13, no. 3, pp. 2213–2227, 2019.

[14] ——, "Data-driven methods for predictive maintenance of industrial equipment: A survey," *IEEE systems journal*, vol. 13, no. 3, pp. 2213–2227, 2019.

[15] Y. Liao, L. Zhang, and C. Liu, "Uncertainty prediction of remaining useful life using long short-term memory network based on bootstrap method," in *2018 ieee international conference on prognostics and health management (icphm)*. IEEE, 2018, pp. 1–8.

[16] S. Zhao, Y. Zhang, S. Wang, B. Zhou, and C. Cheng, "A recurrent neural network approach for remaining useful life prediction utilizing a novel trend features construction method," *Measurement*, vol. 146, pp. 279–288, 2019.

[17] S. Agarwal, V. Saxena, V. Singal, and S. Aggarwal, "Lstm based music generation with dataset preprocessing and reconstruction techniques," in *2018 IEEE symposium series on computational intelligence (SSCI)*. IEEE, 2018, pp. 455–462.

[18] G. Lan, Q. Li, and N. Cheng, "Remaining useful life estimation of turbofan engine using lstm neural networks," in *2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC)*. IEEE, 2018, pp. 1–5.

[19] A. P. Hermawan, D.-S. Kim, and J.-M. Lee, "Predictive maintenance of aircraft engine using deep learning technique," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2020, pp. 1296–1298.

[20] T. Kobayashi and D. L. Simon, "Hybrid kalman filter approach for aircraft engine in-flight diagnostics: sensor fault detection case," in *Turbo Expo: Power for Land, Sea, and Air*, vol. 42371, 2006, pp. 745–755.

[21] S. Chen, M. Wang, D. Huang, P. Wen, S. Wang, and S. Zhao, "Remaining useful life prediction for complex systems with multiple indicators based on particle filter and parameter correlation," *IEEE Access*, vol. 8, pp. 215 145–215 156, 2020.

[22] J. B. Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, "Accurate bearing remaining useful life prediction based on weibull distribution and artificial neural network," *Mechanical Systems and Signal Processing*, vol. 56, pp. 150–172, 2015.

[23] K. Salahshoor, M. Mosallaei, and M. Bayat, "Centralized and decentralized process and sensor fault monitoring using data fusion based on adaptive extended kalman filter algorithm," *Measurement*, vol. 41, no. 10, pp. 1059–1076, 2008.

[24] C. Ordóñez, F. S. Lasheras, J. Roca-Pardiñas, and F. J. de Cos Juez, "A hybrid arima–svm model for the study of the remaining useful life of aircraft engines," *Journal of Computational and Applied Mathematics*, vol. 346, pp. 184–191, 2019.

[25] R. Khelif, B. Chebel-Morello, S. Malinowski, E. Laajili, F. Fnaiech, and N. Zerhouni, "Direct remaining useful life estimation based on support vector regression," *IEEE Transactions on industrial electronics*, vol. 64, no. 3, pp. 2276–2285, 2016.

[26] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2306–2318, 2016.

[27] G. Lan, Q. Li, and N. Cheng, "Remaining useful life estimation of turbofan engine using lstm neural networks," in *2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC)*. IEEE, 2018, pp. 1–5.

[28] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long short-term memory network for remaining useful life estimation," in *2017 IEEE in-*

ternational conference on prognostics and health management (ICPHM). IEEE, 2017, pp. 88–95.

[29] R. Wu and J. Ma, "An improved lstm neural network with uncertainty to predict remaining useful life," in *2019 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*. IEEE, 2019, pp. 274–279.

[30] S. Wang, X. Zhang, D. Gao, B. Chen, Y. Cheng, Y. Yang, W. Yu, Z. Huang, and J. Peng, "A remaining useful life prediction model based on hybrid long-short sequences for engines," in *2018 21st international conference on intelligent transportation systems (ITSC)*. IEEE, 2018, pp. 1757–1762.

[31] G. Sateesh Babu, P. Zhao, and X.-L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life," in *Database Systems for Advanced Applications: 21st International Conference, DASFAA 2016, Dallas, TX, USA, April 16-19, 2016, Proceedings, Part I 21*. Springer, 2016, pp. 214–228.

[32] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, 2018.

[33] L. Jayasinghe, T. Samarasinghe, C. Yuenv, J. C. N. Low, and S. S. Ge, "Temporal convolutional memory networks for remaining useful life estimation of industrial machinery," in *2019 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, 2019, pp. 915–920.

[34] C. W. Hong, C. Lee, K. Lee, M.-S. Ko, D. E. Kim, and K. Hur, "Remaining useful life prognosis for turbofan engine using explainable deep neural networks with dimensionality reduction," *Sensors*, vol. 20, no. 22, p. 6626, 2020.

[35] H. Mo, F. Lucca, J. Malacarne, and G. Iacca, "Multi-head cnn-lstm with prediction error analysis for remaining useful life prediction," in *2020 27th conference of open innovations association (FRUCT)*. IEEE, 2020, pp. 164–171.

[36] Z. Zhao, B. Liang, X. Wang, and W. Lu, "Remaining useful life prediction of aircraft engine based on degradation pattern learning," *Reliability Engineering & System Safety*, vol. 164, pp. 74–83, 2017.

[37] G. Lan, Q. Li, and N. Cheng, "Remaining useful life estimation of turbofan engine using lstm neural networks," in *2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC)*. IEEE, 2018, pp. 1–5.

[38] C. Liu, L. Zhang, R. Yao, and C. Wu, "Dual attention-based temporal convolutional network for fault prognosis under time-varying operating conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[39] S. Sayyad, S. Kumar, A. Bongale, P. Kamat, S. Patil, and K. Kotecha, "Data-driven remaining useful life estimation for milling process: sensors, algorithms, datasets, and future directions," *IEEE Access*, vol. 9, pp. 110 255–110 286, 2021.

[40] A. Essien and C. Giannetti, "A deep learning model for smart manufacturing using convolutional lstm neural network autoencoders," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 6069–6078, 2020.

[41] A. V. Malawade, N. D. Costa, D. Muthirayan, P. P. Khargonekar, and M. A. Al Faruque, "Neuroscience-inspired algorithms for the predictive maintenance of manufacturing systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 7980–7990, 2021.

[42] Y. K. Teoh, S. S. Gill, and A. K. Parlikad, "Iot and fog computing based predictive maintenance model for effective asset management in industry 4.0 using machine learning," *IEEE Internet of Things Journal*, 2021.

[43] K. S. H. Ong, W. Wang, N. Q. Hieu, D. Niyato, and T. Friedrichs, "Predictive maintenance model for iiot-based manufacturing: A transferable deep reinforcement learning approach," *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 15 725–15 741, 2022.