ICCK Transactions on Systems Safety and Reliability
http://dx.doi.org/10.62762/TSSR.2025.167369

RESEARCH ARTICLE

Remaining Useful Life Prediction Using Optimized Multi-source Features and Model Fusion

Xin Zhou[1] **and Lechang Yang**[1,*]

[1] School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China

Abstract

Remaining Useful Life (RUL) prediction is critical for ensuring equipment safety and optimizing maintenance schedules, directly impacting system reliability and maintenance efficiency. However, in real-world industrial scenarios, factors such as operating condition fluctuations and load variations lead to inconsistent data distributions, making it challenging for existing models to achieve satisfactory adaptability and accuracy. To address this issue, this paper proposes a deep learning framework based on a multi-branch serial-parallel fusion of CNN-BiLSTM-Transformer architectures. Through innovative model architecture design and optimized training strategies, the framework aims to enhance prediction accuracy and robustness under cross-condition and cross-distribution scenarios. Specifically, a training strategy combining snapshot ensemble and cosine annealing learning rate scheduling is introduced to improve model generalization through multi-model ensemble prediction. Experiments conducted on the C-MAPSS aero-engine dataset and the A123 lithium battery dataset validate the effectiveness of the proposed method. The results show that on the FD004 dataset, the proposed model reduces RMSE by 7.61% and Score by 26.39% compared to the best baseline methods; on the main test set of the A123 battery dataset, RMSE is reduced by 16.18%. Overall, the proposed model outperforms traditional machine learning models under both single and complex operating conditions, and its prediction results are comparable to those of mainstream models, demonstrating the feasibility and effectiveness of the proposed method in improving RUL prediction accuracy under cross-condition and cross-distribution scenarios.

Keywords: remaining useful life prediction, multi-branch fusion model, aviation engine, lithium battery.

Submitted: 05 August 2025
Accepted: 09 September 2025
Published: 12 November 2025

Vol. 1, **No.** 2, 2025.
10.62762/TSSR.2025.167369

***Corresponding author:**
Lechang Yang
yanglechang@126.com

Citation
Zhou, X., & Yang, L. (2025). Remaining Useful Life Prediction Using Optimized Multi-source Features and Model Fusion. *ICCK Transactions on Systems Safety and Reliability*, 1(2), 114–127.

1 Introduction

In modern industrial systems, RUL prediction for industrial equipment has become a core research topic in predictive maintenance. With increasing complexity and integration of industrial equipment, numerous sensors are deployed to monitor operational states, generating massive multi-source time-series data. These data from diverse industrial equipment exhibit three key common characteristics:

(1) Typically manifest as high-dimensional temporal signals containing rich local patterns and global trends;

(2) Equipment degradation processes often demonstrate nonlinear and non-stationary properties;

(3) Variations in operational conditions significantly influence degradation trajectories.

These shared characteristics provide a critical foundation for developing universal RUL prediction methodologies. Existing RUL prediction approaches fall into three primary categories: physics-based models, data-driven methods, and hybrid approaches. Physics-based models employ physical failure mechanism equations for prediction, e.g., the Paris-Erdogan fatigue crack growth model [1] and windowed linear Kalman filter models [2]. While offering clear physical interpretability, they suffer from complex modeling procedures and limited generalizability. Data-driven methods utilize machine learning algorithms to learn degradation patterns from historical data, encompassing traditional algorithms like Support Vector Regression (SVR) [3] and Random Forests (RF) [4], as well as deep learning models such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN). Hybrid methods attempt to integrate the strengths of both, exemplified by Zhang et al.'s physics-informed neural network framework [5].

However, most existing studies focus on specific equipment types, lacking systematic analysis of commonalities in industrial monitoring data. Furthermore, traditional feature engineering approaches rely heavily on domain expertise, struggling to adapt to complex and variable operating conditions. This results in insufficient model generalizability, where prediction performance degrades significantly under new operational conditions or with new equipment types.

To address these challenges, this paper proposes a CNN-LSTM-Transformer-based deep neural network with hybrid serial-parallel fusion architecture, targeting RUL prediction for industrial equipment exhibiting the aforementioned data characteristics. We further introduce a snapshot ensemble learning strategy [6] with dynamic learning rate scheduling to prevent convergence to local optima. Extensive experiments validate the model's cross-device and cross-data-distribution prediction capabilities on industrial equipment with similar data characteristics. The main contributions are:

(1) Multi-source feature fusion: Integrates automatically extracted deep features with manually constructed statistical features to achieve multi-perspective degradation representation, mitigating limitations of single-data sources.

(2) Novel hybrid serial-parallel fusion architecture: Enables hierarchical fusion of local-global-handcrafted features through collaborative CNN, BiLSTM, Transformer, and handcrafted feature branches, enhancing representation of complex degradation processes.

(3) Cross-device experimental validation: Demonstrates method effectiveness and robustness on C-MAPSS aircraft engine and A123 lithium battery datasets.

The remainder of this paper is organized as follows: Section 2 reviews related RUL prediction studies. Section 3 introduces theoretical foundations and model architecture. Section 4 presents experimental results and analysis on public datasets. Section 5 concludes the work and suggests future research directions.

2 Related Work

With advancements in IoT technology and sensor capabilities, unprecedented data resources have been provided for equipment condition monitoring, making data-driven RUL prediction methods a research hotspot. Such methods no longer rely on specialized knowledge of internal structures or physical failure mechanisms but instead focus on analyzing accumulated operational monitoring data [7]. Among these, neural network-based approaches have emerged as mainstream research directions due to their powerful feature learning capabilities, broadly categorized into single neural networks and hybrid neural networks [8].

For single-network methods, Zhang et al. [9] utilized 1D Convolutional Neural Networks (1D-CNN) for bearing life prediction, demonstrating CNN's effectiveness in processing vibration signals. In hybrid networks, Li et al. [10] developed a Deep Convolutional Neural Network (DCNN) capable of directly learning degradation features from raw vibration data, reducing dependence on traditional signal processing expertise; Zhang et al. [11] employed supervised LSTM networks to capture long-term temporal dependencies, enabling effective modeling of complex degradation processes. Miao et al. [12] proposed a dual-task deep LSTM architecture that simultaneously predicts RUL and health status, enhancing overall model performance.

Integrating physical models with data-driven methods can significantly improve model adaptability and prediction accuracy. Hybrid approaches primarily adopt two technical pathways:

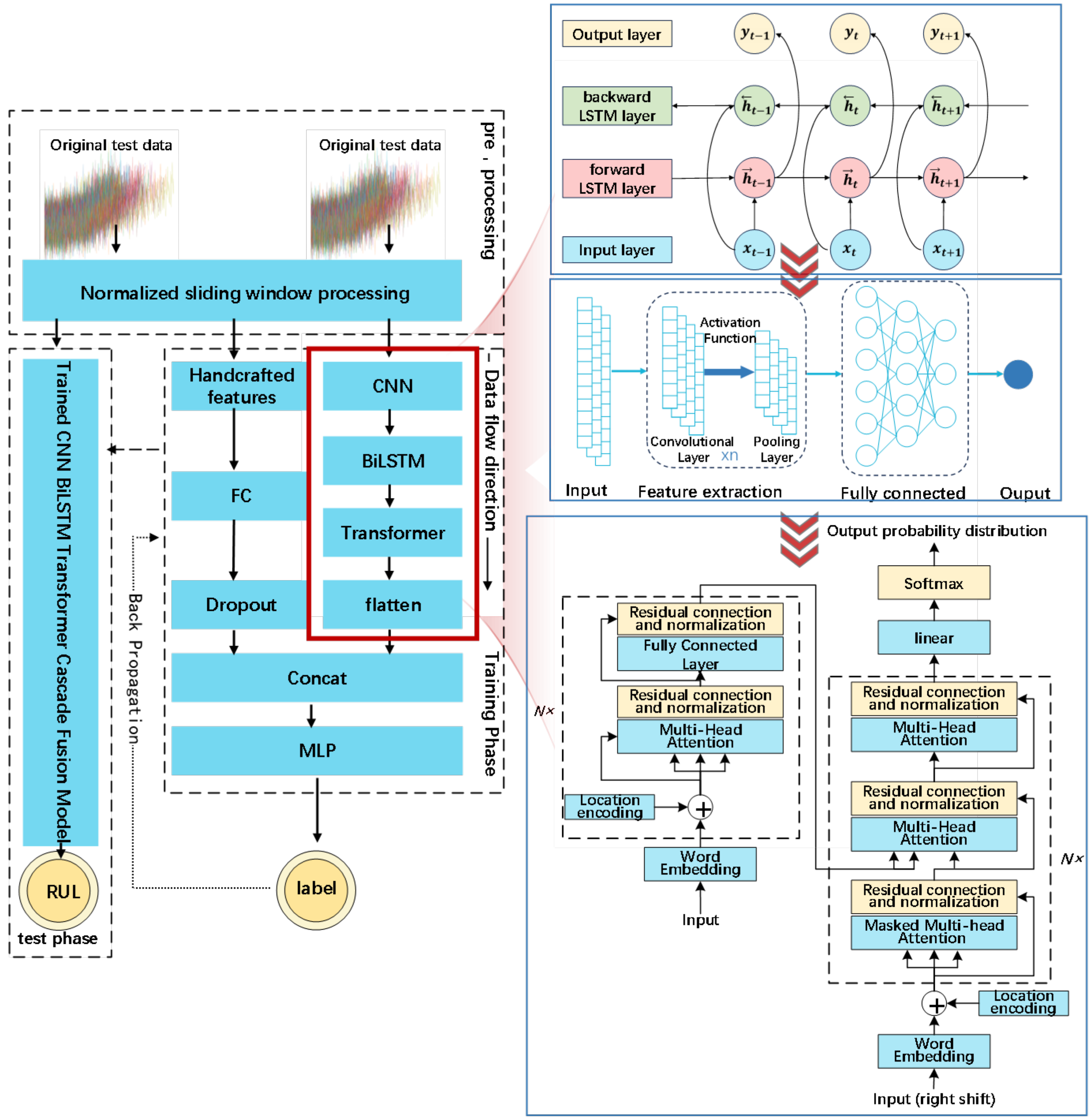


Figure 1. Structure of the model.

(1) Physics-model and machine learning synergy: Physical models describe degradation mechanisms while machine learning algorithms handle uncertainties.

(2) Multi-model ensemble strategies: Integrating predictions from diverse models to construct hierarchical frameworks.

Specifically, base models generate preliminary predictions, which are then fed as new features into meta-models for secondary learning. Current trends indicate data-driven methods dominate due to limitations of physics-based approaches in engineering practice [13].

Rapid developments in deep learning have enabled hybrid neural architectures to demonstrate significant advantages in RUL prediction, particularly in capturing multi-scale temporal features:

CNN effectively extracts local patterns from time-series data, while LSTM networks [14] and their variants model long-term degradation trends. Transformer architectures, on the other hand, excel at capturing long-range dependencies in sequential data. Building on these capabilities, researchers have explored various hybrid architectures to enhance prognostics performance. In CNN-LSTM hybrids, Ma et al. [15] integrated CNN-based feature extraction with LSTM for temporal modeling in lithium-ion battery remaining useful life (RUL) prediction. Similarly, Ren et al. [16] introduced an autoencoder-enhanced CNN-LSTM framework to boost feature dimensionality and extract deeper degradation insights for battery RUL. For Transformer-based approaches, innovations such as data-denoised improved Transformers [17] have demonstrated notable accuracy gains in battery prognostics. Liu et al. [18] developed a dual-attention mechanism fusing CNN and Transformer elements with channel-temporal attention for aircraft engine RUL estimation. Li et al. [19] advanced aero-engine RUL prediction by combining variational mode decomposition, K-means clustering, and dual-channel Transformers with knowledge-driven deep learning. Kang et al. [20] applied artificial neural networks, including multi-model fusions like CNN-LSTM-Attention and Transformer-LSTM variants augmented by attention and feature reconstruction, to predict equipment RUL in production lines under diverse conditions. Specifically for lithium-ion batteries, Chen et al. [21] introduced a hybrid channel attention-LSTM (CA-LSTM) model that dynamically weights feature channels, mitigates capacity rebound noise, and captures temporal degradation patterns, yielding robust results across multiple starting points in NASA and CALCE datasets. Despite these advances, current methods continue to face challenges in noise mitigation and multi-feature integration, especially for ensuring reliability with noisy, real-world operational data.

3 Method

3.1 Model Design

While architectures combining CNNs, RNNs, and Transformers exist, they are often designed to maximize accuracy on a single dataset or operational condition. The key novelty of our work lies in the deliberate design of a fusion architecture for cross-distribution robustness. To this end, we introduce a novel hierarchical serial-parallel structure that is explicitly engineered to capture both complex temporal patterns and stabilizing domain-knowledge features, which is crucial for generalizing to new conditions. Our core innovation is the novel hierarchical fusion pathway: a serial-progressive branch (CNN → BiLSTM → Transformer) that systematically abstracts features from local to global temporal contexts, coupled in parallel with a handcrafted feature branch. This specific co-design is a distinct architectural strategy for learning robust, domain-invariant representations, moving beyond the capabilities of sequential or simpler parallel hybrids. As illustrated in Figure 1, the model input comprises two components: preprocessed multi-dimensional time-series sensor data and domain-knowledge-driven handcrafted features.

The proposed model processes time-series data through a structured pipeline designed for comprehensive feature extraction and fusion. In the temporal processing branch, raw sensor data first undergoes local feature extraction via a 1D convolutional layer with kernel size 3 and stride 1, utilizing ReLU activation functions to capture short-term dependencies and localized patterns within sliding windows. The resulting feature maps are then fed into a BiLSTM module with 64 hidden units per direction. This module processes sequences in both forward and backward directions, effectively modeling long-term temporal dependencies through its gated memory mechanism and hidden state propagation across time steps. To further enhance global temporal modeling, a Transformer encoder layer is incorporated following the BiLSTM module. This component employs multi-head self-attention with 4 parallel attention heads, each operating in distinct feature subspaces to simultaneously capture diverse dependency patterns between any two time steps regardless of their positional distance. The attention outputs are processed through position-wise feedforward networks with 128-dimensional hidden layers and residual connections, followed by layer normalization to stabilize training. This sequential CNN-BiLSTM-Transformer architecture enables progressive feature abstraction-from low-level local patterns to mid-range dependencies and finally to global temporal relationships-effectively overcoming the limitations of individual architectures in capturing complex temporal dynamics.

Simultaneously, a parallel handcrafted feature branch processes domain-specific statistical features including mean values, trend coefficients, and covariance measures. These features undergo transformation through two fully connected layers with ReLU activation functions and Dropout regularization after each layer to prevent overfitting and enhance robustness. This branch performs nonlinear feature transformation and dimensionality reduction while preserving critical domain knowledge.

During feature fusion, the temporal features from the main branch and the processed handcrafted features are concatenated along the feature dimension to form a unified representation. The combined feature vector subsequently passes through a regression network consisting of two fully connected layers with ReLU activations, followed by a final linear output layer that produces the RUL prediction. This integrated approach enables simultaneous learning from both data-driven temporal patterns and expert knowledge, significantly enhancing the model's representational capacity and generalization performance across varying operational conditions.

3.2 Feature Extraction

The quality of feature engineering critically determines model prediction performance. This study systematically extracts and constructs two key feature categories from sensor time-series data: domain-knowledge-driven handcrafted features and raw sensor features. For handcrafted features, we extract multiple statistical indicators reflecting equipment degradation characteristics, including mean values, linear regression coefficients, and covariance between sensor data and actual lifespan. These features reveal performance degradation patterns across different dimensions, enhancing the model's ability to learn complex degradation modes. Regarding raw sensor feature selection, we employ Pearson and Spearman correlation coefficients to screen raw monitoring data, with the Spearman coefficient calculated as follows (Eq. 1):

$$\rho\left(X,Y\right) = 1 - \frac{6\sum_{i=1}^{n}\left(R(X_i) - R(Y_i)\right)}{n(n^2-1)} \tag{1}$$

where $R(X_i)$ and $R(Y_i)$ denote the ranks of data points in vectors X and Y respectively, and n represents the sample size.

Primary feature subsets exhibiting correlations beyond this predetermined threshold are selected through this analysis. Subsequently, Recursive Feature Elimination (RFE) is performed using an Extreme Gradient Boosting Regressor (XGBRegressor) with 100 estimators and maximum depth of 6 on these subsets. The RFE process iteratively removes the least important features based on feature importance rankings until optimal feature subset size is achieved. This two-stage approach ultimately yields discriminative sensor feature subsets that serve as raw inputs to the model, ensuring both relevance and minimal redundancy in the input feature space.

3.3 Evaluation Metrics

Three widely accepted evaluation metrics are adopted to assess model performance: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Scoring metric. The prediction Score-a direction-sensitive asymmetric evaluation criterion established by NASA -imposes stricter penalties on delayed predictions, effectively reflecting practical consequences of different prediction deviations. Its mathematical formulation is defined in Equation (2):

$$s = \begin{cases} \sum_{i=1}^{n} e^{-\frac{y_i'-y_i}{13}} - 1, y_i' - y_i < 0 \\ \\ \sum_{i=1}^{n} e^{-\frac{y_i'-y_i}{10}} - 1, y_i' - y_i > 0 \end{cases} \tag{2}$$

where y_i' denotes the prediction error for the i-th sample, with y_i representing the predicted and actual Remaining Useful Life values respectively, and n indicates the total number of samples.

4 Experiments

4.1 Aviation Engine Experiment

4.1.1 Dataset

The NASA C-MAPSS dataset comprises four sub-datasets, with primary distinctions lying in their operational condition complexity and fault mode combinations. Each data record includes real-time readings from 21 sensors along with three operational condition parameters. To validate the adaptability and robustness of the proposed model across diverse application scenarios, experiments were conducted on both the FD001 sub-dataset (representing single operational conditions) and the FD004 sub-dataset (characterizing complex operational conditions).

4.1.2 Data Preprocessing

(1) Sensor Parameter Selection and Normalization. Using the FD001 sub-dataset of the NASA C-MAPSS dataset as a case study, only 14 out of 21 sensors demonstrate highly correlated dynamic variations with RUL throughout the full degradation cycle

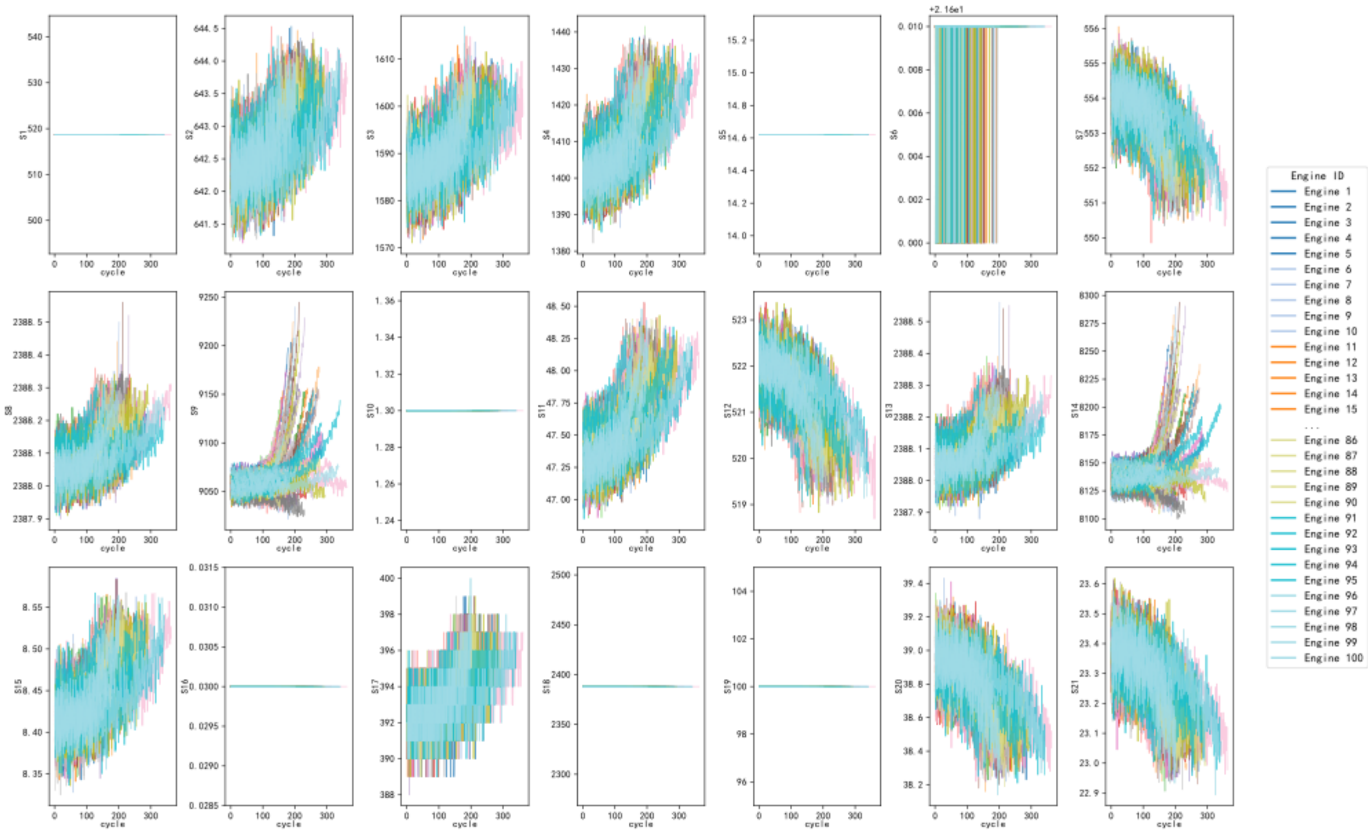


Figure 2. Information of 21 sensors in the FD001 training set.

(Figure 2). Given significant disparities in units and value ranges across sensors, an enhanced Min-Max normalization method is applied for data standardization.

(2) Sliding Window and Label Setting. In this study, a sliding window approach is employed to sample the engine data, with a window size of 30 and a stride of 1 to generate the model input data. When the engine is in its early healthy state, the overall degradation trend remains stable and gradual, and the RUL can be considered constant during this phase. Therefore, a piecewise RUL labeling strategy is adopted, where the RUL label is set to 125 cycles.

$$RUL_n^m = \min \{h^m - h_n^m, 125\} \tag{3}$$

where RUL_n^m denotes the RUL label of the n-th time window for the m-th engine unit, h^m represents the total operational cycles over the full degradation life of the m-th engine, and h_n^m indicates the terminal cycle index of the n-th time window for the m-th engine.

4.1.3 Feature Selection

Handcrafted features-including mean values, trend coefficients, covariance, and other statistical indicators-characterize equipment degradation patterns from multiple dimensions. Through Pearson correlation analysis, 14 key parameters were identified as the raw sensor feature set.

4.1.4 Experimental Parameter Settings

In this study, mean squared error (MSE) is adopted as the loss function. The model architecture integrates a CNN, a BiLSTM network, a Transformer encoder, and a handcrafted feature branch to fully exploit the multi-level features of time-series data. The output features from each branch are concatenated and then fed into a fully connected network for final prediction. Detailed network parameter configurations are provided in Table 1.

For model training, the Adam optimizer is employed with dynamic learning rate adjustment. The maximum number of training epochs is set to 32, and the remaining parameter configurations are summarized in Table 2.

4.1.5 Analysis of Experimental Results

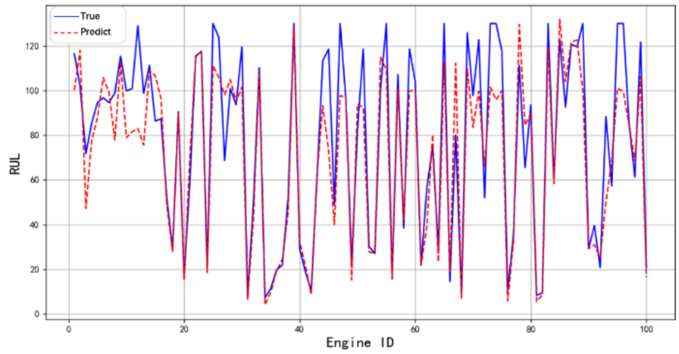
(1) Overall performance evaluation. In this section, the model is trained on the C-MAPSS dataset. Figure 3(a) and 3(b) show the prediction results of the model on the FD001 and FD004 test sets, respectively.

Table 1. Aero-engine network architecture parameters.

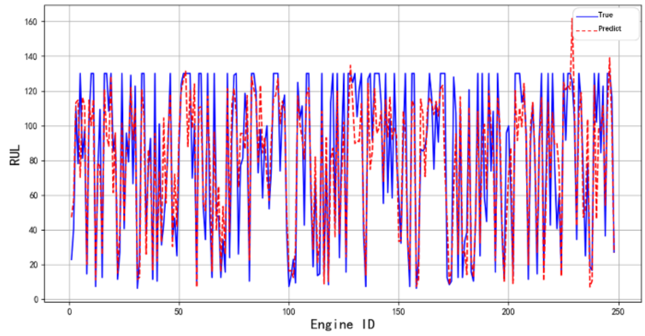
Network Layer	Parameters
Conv1d	Kernel size:3
BiLSTM	Hidden size:64
Transformer Encoder	Num_heads: 4; The number of nodes in the first layer of the feedforward network: 128 Dropout: 0.2
FC1	Dropout:0.2
FC2	Dropout:0.2

Table 2. Training hyperparameters for aero-engine experiments.

Parameter	Description	Value
iterations	Number of training repetitions	5
Window size	Time window size	30
lr_1	Learning rate	0.005
R_seed_list	Random seed list	[42, 123, 456, 789, 999]
Epochs	Number of training rounds	32
Patience	Early stopping parameters	20
Batch size	Batch size	256



(a)Dataset:FD001



(b)Dataset:FD004

Figure 3. Overall performance evaluation.

Overall, the predicted curves on both the FD001 and FD004 datasets closely fit the actual RUL variation trends. In particular, during the late stage of engine life degradation, the predicted trajectories exhibit a high degree of consistency with the true degradation curves. This indicates that the proposed model can effectively capture the key pattern features in the engine degradation process.

(2) Comparative experimental results. Several mainstream traditional machine learning methods and deep learning methods are selected as baseline models for comparison. The detailed comparison results are presented in Table 3.

As shown in the table, the proposed model achieves an RMSE of 19.66 and a Score of 1976.3 on the FD004 dataset. Compared with the current best baseline method, the RMSE and Score of the proposed model on FD004 are reduced by 7.61% and 26.39%, respectively, demonstrating strong generalization ability and robustness. It is worth noting that on the FD001 dataset, while the model's RMSE (14.02) and Score (284.17) are slightly higher than those of some methods highly optimized for single-condition scenarios (such as references [25] and [26]), this is due to a deliberate trade-off: our model prioritizes learning domain-invariant feature representations to ensure cross-condition robustness, which may marginally reduce its optimization for a single, stable operating condition. In contrast, the proposed model excels significantly on the more challenging FD004 subset, which involves multiple operating conditions and fault modes. Moreover, it achieves the best average RMSE and average Score across both the FD001 and FD004 datasets among all compared models. All experiments were conducted on an NVIDIA GeForce RTX 3050 Laptop GPU. The proposed model requires approximately 30 seconds for a single run. These results fully demonstrate that the proposed model not only adapts well to prediction tasks under single operating conditions, but also maintains high prediction accuracy and stability under complex and variable conditions, confirming the superior overall performance and robustness of the CNN-BiLSTM-Transformer model in aero-engine RUL prediction.

The performance results of the baseline methods are directly obtained from their original publications [5, 22–26] to ensure a fair and consistent comparison with

Table 3. RUL prediction performance comparison on C-MAPSS datasets.

Methods	FD001		FD004		Average	
	RMSE	Score	RMSE	Score	RMSE	Score
MLP [22]	37.56	17972	77.37	5616600	57.47	2817286
SVR [22]	20.96	1381.5	45.35	371140	33.16	186260.8
RVR [22]	23.80	1502.9	34.34	26509	29.07	14005.95
ELM [5]	17.27	523.00	38.43	121414.47	27.85	60968.74
RF [5]	17.91	479.75	31.12	46567.63	24.52	23523.69
CNN [22]	18.45	1286.7	29.16	7886.4	23.81	4586.55
LSTM [23]	15.42	410.60	29.12	12551.44	22.27	6481.02
DBN [5]	15.21	417.59	29.88	7954.51	22.55	4186.05
MODBNE [5]	15.04	334.23	28.66	6557.62	21.85	3445.925
BiLSTM [24]	13.65	295	27.08	5430	20.37	2862.5
MSDCNN-LSTM [25]	13.19	247.69	21.28	2684.65	17.24	1466.17
RCNN ABiLSTM [26]	12.98	258	22.29	3795	17.64	2026.5
Ours	14.02	284.17	19.66	1976.3	16.84	1130.235

state-of-the-art methods under their reported optimal conditions. To maximize comparability, we have strictly adhered to the standard dataset splits (FD001 and FD004), evaluation metrics (RMSE and Score), and RUL clipping strategies commonly adopted in these studies. Although direct reimplementation was not feasible due to constraints in code availability and reproducibility, we acknowledge that differences may exist in data preprocessing-such as normalization techniques, sliding window configurations, or feature engineering approaches-which could lead to minor variations in absolute performance. Nevertheless, the substantial and consistent improvements of our method across both single and complex operational conditions (e.g., 7.61% RMSE reduction on FD004 and superior average performance) robustly demonstrate its effectiveness and generalizability.

4.2 Lithium Battery Experiment

4.2.1 Dataset

The "A123 dataset" used in this study was collected by Severson et al. [27] and contains 124 commercial lithium-ion battery cells. This dataset records cycling test data of batteries under various fast-charging protocols; all cells were discharged at a uniform rate until reaching end of life (i.e., when capacity decayed to 80% of the nominal value). The recorded performance parameters include voltage, current, charge/discharge capacity, temperature, internal resistance, and charging duration.

The dataset comprises battery cells from three different batches. Cells from the first two batches are alternately divided into the training set and the primary test set, while all cells from the third batch are assigned to the secondary test set. It is noteworthy that the training and primary test sets were collected during the same period, whereas the secondary test set was collected approximately one year later. As a result, the secondary test set is affected by factors such as calendar aging, leading to a data distribution shift compared to the training and primary test sets. As shown in Figure 4, the figure compares the distribution characteristics of battery cells in the three subsets in terms of cycle life and initial discharge capacity. It can be observed that the distributions of the training and primary test sets are relatively similar, while the secondary test set exhibits significant differences.

Based on this dataset, this section evaluates the generalization ability of the proposed model on both the primary and secondary test sets. Detailed information about the A123 battery dataset is available at: https://data.matr.io/1.

4.2.2 Data Preprocessing

In the A123 battery dataset, the capacity degradation trajectories of individual cells are shown in Figure 5. It can be observed that when the cycle number reaches approximately 400, most cells exhibit a pronounced inflection point in their capacity curves. After this inflection point, the battery capacity enters a rapid decay phase, displaying an accelerated degradation trend. Additionally, some data points are identified as clear outliers, which may be attributed to occasional disturbances during data acquisition or experimental operations. After data cleaning and filtering, valid

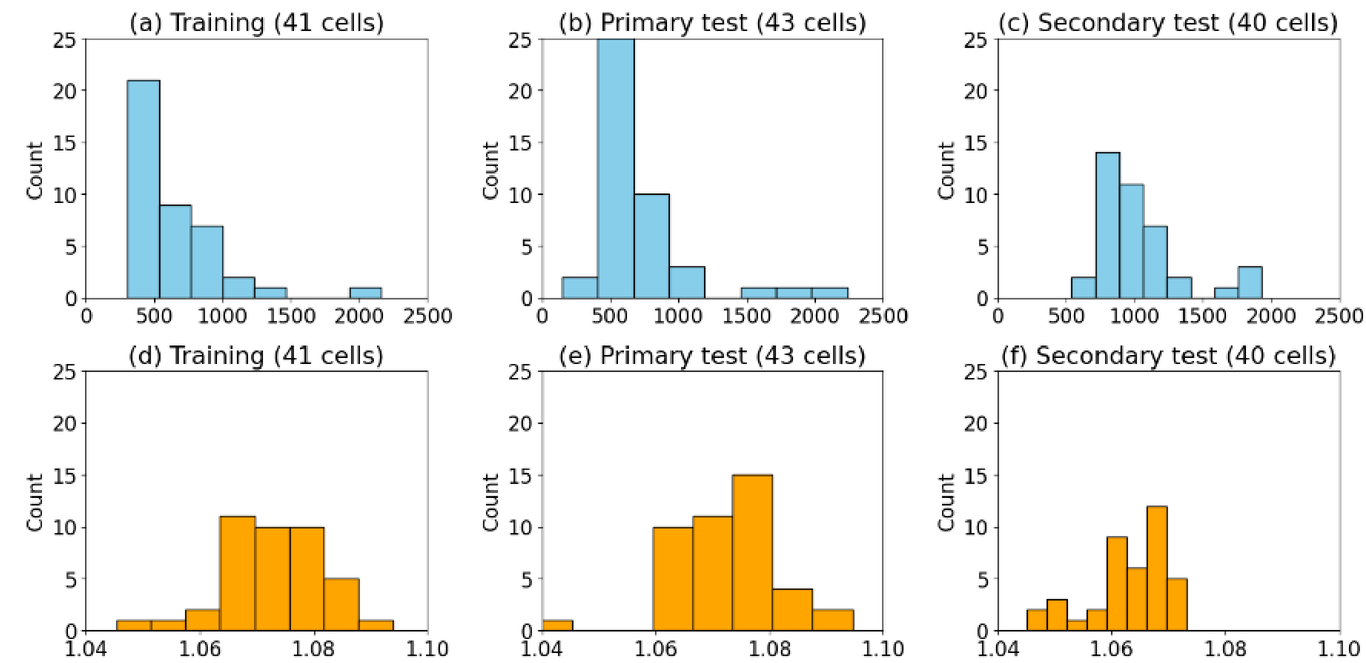


Figure 4. (a)-(c) shows the distribution of battery groups for training, main test, and secondary test based on cycle life; (d)-(f) Distribution of battery groups for training, main test, and secondary test based on the initial capacity.

data from 124 battery cells were retained.

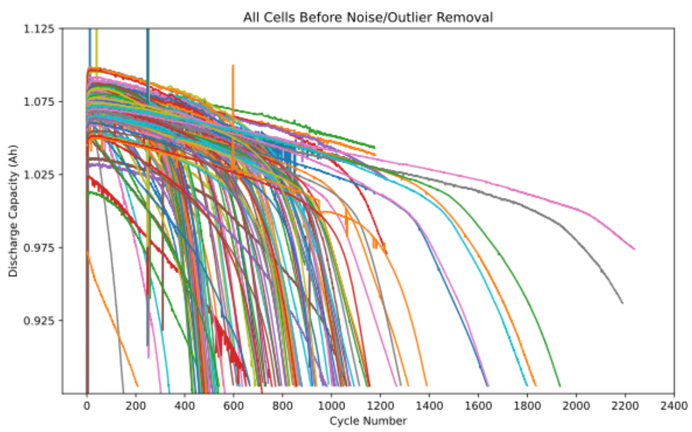


Figure 5. Battery capacity degradation curve.

4.2.3 Feature Extraction and Correlation Analysis

(1) Domain Feature Extraction and Correlation Analysis. Severson et al. [27] proposed 20 domain-knowledge-driven features for the A123 battery dataset and analyzed their correlations with battery RUL. For instance, the variance of feature $\Delta Q_{100-10}(V)$ (var_dQ_100_10) exhibits a Spearman correlation coefficient of 0.93 with RUL, indicating a strong monotonic relationship. This single feature alone enables high RUL prediction accuracy. However, such expert-derived features are highly dependent on domain expertise, inherently subjective, and sensitive to data distribution shifts, limiting their generalization capability. Consequently, integrating domain knowledge with data-driven feature optimization is crucial for enhancing the robustness of RUL prediction.

In this study, we employed the Spearman correlation coefficient (threshold > 0.45) combined with XGBRegressor-based Recursive Feature Elimination to identify eight key features exhibiting strong correlations with RUL for modeling: var_dQ_100_10, min_dQ_100_10, mean_dQ_100_10, slope_Qd_cycle_91_100, Qd_100,min_ir_cycle_2_100, ir_diff_cycle_100_2, and Qd_2. This specific threshold value was determined through grid search optimization on the validation set, representing an optimal balance between feature selectivity and information preservation. This threshold ensures that only features with substantial statistical association with degradation progression are retained while maintaining adequate feature diversity for cross-condition generalization.

(2) Handcrafted features Feature Extraction. To thoroughly explore the effective information embedded in battery cycling data, this study extracts diverse statistical features from key variables including discharge capacity, temperature, and differential capacity. In contrast to feature engineering approaches reliant on domain expertise, this handcrafted feature extraction method based on statistical metrics demonstrates minimal reliance on prior knowledge,

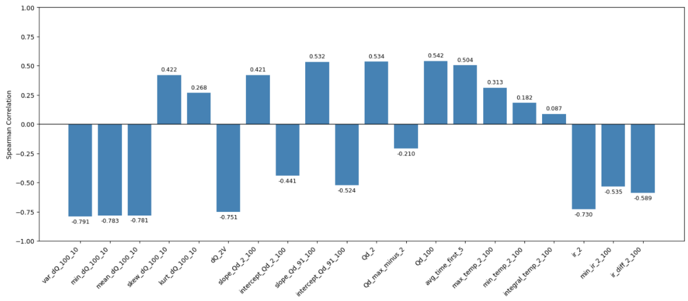


Figure 6. Spearman correlation coefficients between training features and targets.

offering enhanced versatility and scalability.

4.2.4 *Model Hyperparameter Configuration*

To optimize the hyperparameter settings of the hybrid model, this study employed a Bayesian optimization-based automated tuning strategy to perform a global search within a predefined hyperparameter space. Following each parameter sampling iteration, the model was trained on the full training set, with the RMSE on the validation set serving as the objective function for optimization. The hyperparameter configuration yielding optimal performance on the validation set is detailed in Table 4.

Table 4. Optimized hyperparameters for lithium battery model.

Parameter	Description	Range	Value
hidden_dim	Hidden layer dimension	[32,64,128,256]	256
kernel_size	Kernel size	[3,7]	5
lr_2	Learning rate	[1e-5,1e-3]	1e-5

To enhance the robustness and generalization capability of the proposed model, we employed a snapshot ensemble strategy based on cosine annealing learning rates. During the training process, the learning rate periodically varied between the maximum value (0.1) and minimum value (1e-6) following a cosine function, with each cycle spanning 200 training epochs. At the end of each cycle, a model snapshot was automatically saved, and the learning rate was reset to its initial value to restart the decay process. The training parameter configuration is shown in Table 5.

For the final ensemble construction, we selected the last 10 model snapshots generated during training. The ensemble prediction was obtained by averaging the outputs of these 10 snapshots using equal weights (arithmetic mean). This approach ensures that the ensemble incorporates diverse model parameters from different phases of the training process, each potentially occupying different regions in the loss landscape. The equal weighting strategy was adopted to maintain simplicity and avoid introducing additional hyperparameters for weight optimization, while still effectively capturing the benefits of model diversity.

Table 5. Snapshot ensemble training parameters for lithium batteries.

Parameter	Description	Value
iterations	Number of training repetitions	5
R_seed_list	Random seed list	[42, 123, 456, 789, 999]
transformer_head	Number of Transformer heads	8
batch_size	Batch size	64
epoch	Number of training rounds	4000
T_max	Cosine annealing restart period	200
T_mult	Multiplier for restart periods	1

4.2.5 *Analysis of Experimental Results*

This section evaluates the performance of the proposed model on the A123 battery dataset. Experimental results demonstrate that the model achieves exceptional performance on both the primary and secondary test sets. As illustrated in Figure 7, the predicted RUL curves closely track the ground-truth values, indicating that the model effectively captures critical degradation characteristics in lithium-ion batteries and delivers accurate RUL predictions for the vast majority of test samples.

To further validate the efficacy of the proposed CNN-BiLSTM-Transformer model for lithium-ion battery RUL prediction, we conducted a comparative analysis against the Hybrid Ensemble architecture. The results of this comparison are presented in Figure 8. From the figure, it can be observed that the method proposed in this paper achieves relatively good performance on both the main test set and the secondary test set compared to the methods being compared.

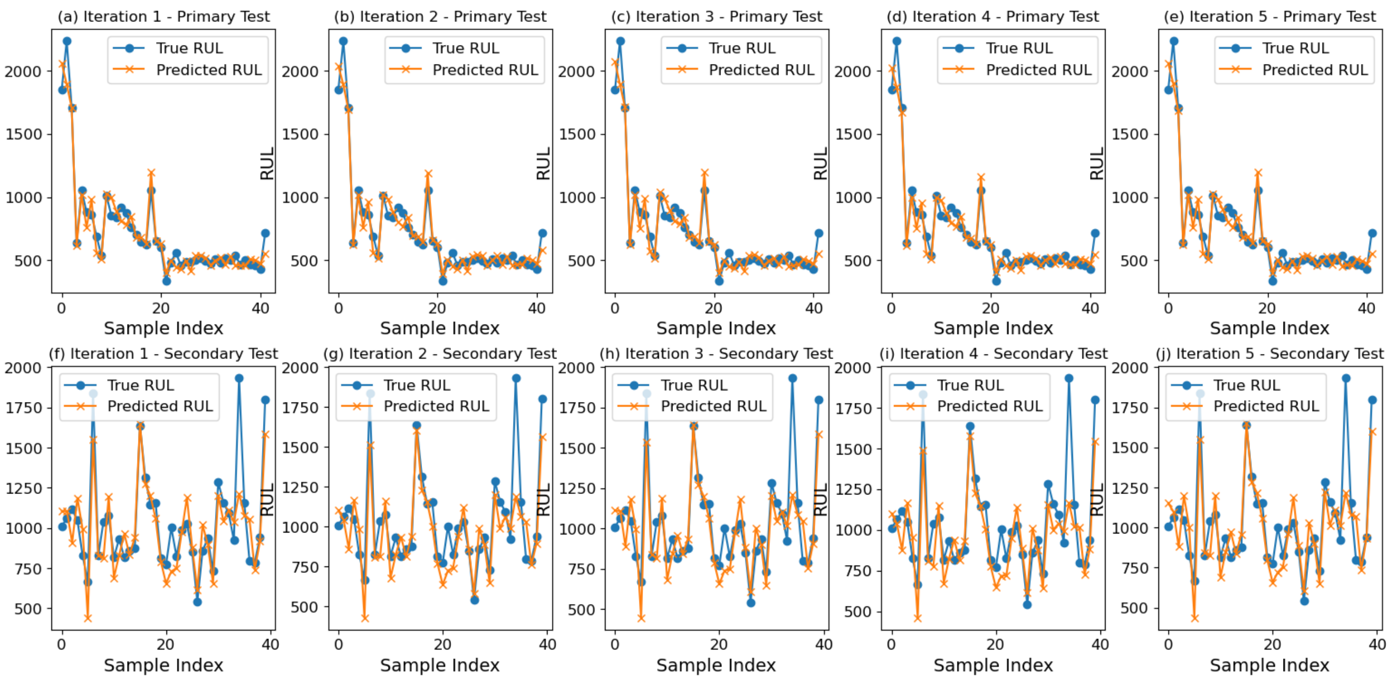


Figure 7. RUL prediction results of the main and secondary test sets during the 5rd iteration process.

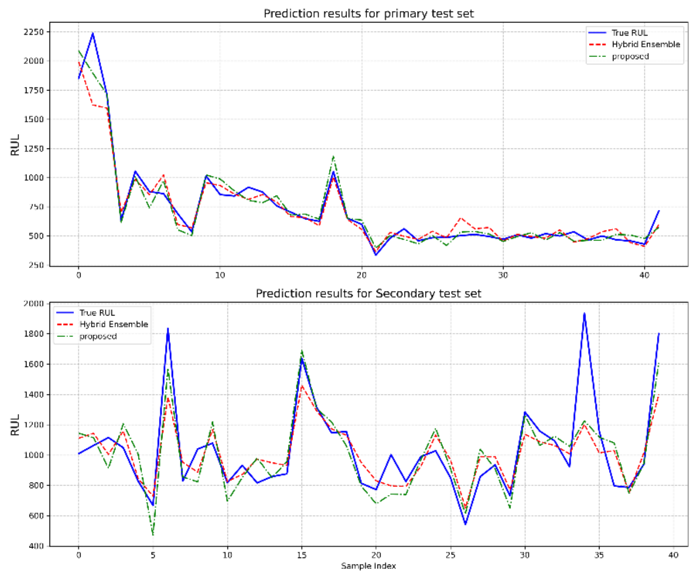


Figure 8. Comparison of RUL prediction results on the main test set and the secondary test set.

Additionally, several mainstream methods were selected as comparative baselines. The models were evaluated using RMSE and MAPE as primary performance metrics, with detailed comparative results presented in Table 6.

Analysis of Table 6 demonstrates the strong performance of the proposed model on the A123 battery dataset. On the primary test set, it achieved RMSE and MAPE values of 95.60 and 8.74, respectively. Compared to the current state-of-the-art baseline model, this represents a 16.18% reduction in RMSE, despite a marginal 2.34% increase in MAPE. On the secondary test set, the model attained RMSE and MAPE values of 187.71 and 12.61, outperforming most traditional machine learning methods and several deep learning approaches, indicating robust comprehensive performance. In terms of computational efficiency, a single run of the proposed model takes approximately 60 seconds. Although the architecture incorporates a multi-branch design, the overall computational cost remains manageable due to the use of a limited number of attention heads and, more importantly, the shorter input sequence length enforced by the sliding window strategy. This approach effectively balances prediction accuracy and computational efficiency.

Furthermore, the proposed model achieved the lowest average RMSE (141.655) across both primary and secondary test sets among all compared methods. Although its average MAPE was marginally higher than that of the Hybrid Ensemble model, it remained superior to most competitors. These results collectively demonstrate the proposed model's exceptional performance and robust generalization capability for lithium-ion battery RUL prediction tasks.

The results for the lithium-ion battery baseline models are taken from the cited references [6, 27–31] to facilitate direct benchmarking. All methods were evaluated using the same primary and secondary

Table 6. Performance comparison with other methods.

Methods	Primary test set		Secondary test set		Average	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Elastic-V [27]	138.39	13.19	196.01	11.41	167.2	12.3
Elastic-D [27]	170.35	10.99	179.64	14.20	174.995	12.595
Elastic-F [27]	117.64	9.20	225.72	12.85	171.68	11.025
SVR-V [28]	170.72	14.72	226.79	12.11	198.755	13.415
NN-F [29]	116.54	9.13	225.83	12.81	171.185	10.97
LSTM [30]	166.87	18.05	380.85	22.66	273.86	20.355
CNN-LSTM [15]	176.63	14.85	375.47	25.25	276.05	20.05
GRU-RNN [16]	127.65	9.94	356.31	34.17	241.98	22.055
PSR-SVR [31]	191.83	20.3	404.57	32.93	298.2	26.615
Hybrid Ensemble [6]	114.05	8.54	177.88	11.31	145.965	9.925
Ours	95.60	8.74	187.71	12.61	141.655	10.675

test sets defined in [27], under consistent evaluation protocols. While differences in preprocessing-such as outlier handling, capacity interpolation, feature scaling, or cycle alignment-might influence numerical outcomes, the significant outperformance of our model, particularly under distribution-shifted scenarios (e.g., 16.18% RMSE reduction on the primary test set and leading average performance), underscores its robustness and advanced generalization capability.

5 Conclusions

This paper proposes a novel multi-branch deep learning framework that integrates 1D CNN, BiLSTM, Transformer, and handcrafted features for robust remaining useful life prediction under cross-domain distribution shifts. The architecture effectively combines local feature extraction, sequential dependency modeling, and long-range pattern capture, significantly enhancing prediction accuracy and generalization across diverse industrial applications. Experimental validation demonstrates strong performance in both aerospace and energy storage domains. On the C-MAPSS dataset, the model achieves competitive results with notable improvements in late-stage predictions, making it suitable for aircraft engine health monitoring and predictive maintenance. For lithium-ion batteries, the model shows excellent generalization on the A123 dataset, enabling accurate battery health management in electric vehicles and grid-scale energy storage systems.

The framework also presents broad applicability in manufacturing and energy sectors for monitoring critical equipment such as turbines, pumps, and power transmission systems. Its adaptability to varying operational conditions and multi-source data makes it particularly valuable for Industrial IoT implementations. Future work will focus on integrating few-shot learning for extreme conditions and combining data-driven with physics-based methods to enhance interpretability and domain adaptation.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Liao, L. (2013). Discovering prognostic features using genetic programming in remaining useful life prediction. *IEEE Transactions on Industrial Electronics, 61*(5), 2464-2472. [Crossref]

[2] Orchard, M. E., & Vachtsevanos, G. J. (2007, June). A particle filtering-based framework for real-time fault diagnosis and failure prognosis in a turbine engine. In *2007 mediterranean conference on control & automation* (pp. 1-6). IEEE. [Crossref]

[3] Anand, P., Rastogi, R., & Chandra, S. (2020). A class of new support vector regression models. *Applied Soft Computing, 94*, 106446. [Crossref]

[4] Mawonou, K. S., Eddahech, A., Dumur, D., Beauvois, D., & Godoy, E. (2021). State-of-health estimators coupled to a random forest approach for lithium-ion battery aging factor ranking. *Journal of Power Sources, 484*, 229154. [Crossref]

[5] Zhang, C., Lim, P., Qin, A. K., & Tan, K. C. (2016). Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE transactions on neural networks and learning systems, 28*(10), 2306-2318. [Crossref]

[6] Xu, Q., Wu, M., Khoo, E., Chen, Z., & Li, X. (2023). A hybrid ensemble deep learning approach for early prediction of battery remaining useful life. *IEEE/CAA Journal of Automatica Sinica, 10*(1), 177-187. [Crossref]

[7] Shifat, T. A., Yasmin, R., & Hur, J. W. (2021). A data driven RUL estimation framework of electric motor using deep electrical feature learning from current harmonics and apparent power. *Energies, 14*(11), 3156. [Crossref]

[8] Zhang, J., Tian, J., Li, M., Leon, J. I., Franquelo, L. G., Luo, H., & Yin, S. (2022). A parallel hybrid neural network with integration of spatial and temporal features for remaining useful life prediction in prognostics. *IEEE Transactions on Instrumentation and Measurement, 72*, 1-12. [Crossref]

[9] Zhang, Q., Ye, Z., Shao, S., Niu, T., & Zhao, Y. (2022). Remaining useful life prediction of rolling bearings based on convolutional recurrent attention network. *Assembly Automation, 42*(3), 372-387. [Crossref]

[10] Li, X., Ding, Q., & Sun, J. Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety, 172*, 1-11. [Crossref]

[11] Zhang, Y., Hutchinson, P., Lieven, N. A., & Nunez-Yanez, J. (2020). Remaining useful life estimation using long short-term memory neural networks and deep fusion. *IEEE access, 8*, 19033-19045. [Crossref]

[12] Miao, H., Li, B., Sun, C., & Liu, J. (2019). Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks. *IEEE Transactions on Industrial Informatics, 15*(9), 5023-5032. [Crossref]

[13] Rezamand, M., Kordestani, M., Orchard, M. E., Carriveau, R., Ting, D. S. K., & Saif, M. (2020). Improved remaining useful life estimation of wind turbine drivetrain bearings under varying operating conditions. *IEEE Transactions on Industrial Informatics, 17*(3), 1742-1752. [Crossref]

[14] Zhang, Y., Xiong, R., He, H., & Pecht, M. G. (2018). Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Transactions on Vehicular Technology, 67*(7), 5695-5705. [Crossref]

[15] Ma, G., Zhang, Y., Cheng, C., Zhou, B., Hu, P., & Yuan, Y. (2019). Remaining useful life prediction of lithium-ion batteries based on false nearest neighbors and a hybrid neural network. *Applied Energy, 253*, 113626. [Crossref]

[16] Ren, L., Dong, J., Wang, X., Meng, Z., Zhao, L., & Deen, M. J. (2020). A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. *IEEE transactions on industrial informatics, 17*(5), 3478-3487. [Crossref]

[17] Zhou, K., & Zhang, Z. (2024). Remaining useful life prediction of lithium-ion batteries based on data denoising and improved transformer. *Journal of Energy Storage, 100*, 113749. [Crossref]

[18] Liu, L., Song, X., & Zhou, Z. (2022). Aircraft engine remaining useful life estimation via a double attention-based data-driven architecture. *Reliability Engineering & System Safety, 221*, 108330. [Crossref]

[19] Li, Y., Chen, Y., Hu, Z., & Zhang, H. (2023). Remaining useful life prediction of aero-engine enabled by fusing knowledge and deep learning models. *Reliability Engineering & System Safety, 229*, 108869. [Crossref]

[20] Kang, Z., Catal, C., & Tekinerdogan, B. (2021). Remaining useful life (RUL) prediction of equipment in production lines using artificial neural networks. *Sensors, 21*(3), 932. [Crossref]

[21] Chen, C., Wei, J., & Li, Z. (2023). Remaining useful life prediction for lithium-ion batteries based on a hybrid deep learning model. *Processes, 11*(8), 2333. [Crossref]

[22] Sateesh Babu, G., Zhao, P., & Li, X. L. (2016, March). Deep convolutional neural network based regression approach for estimation of remaining useful life. In *International conference on database systems for advanced applications* (pp. 214-228). Cham: Springer International Publishing. [Crossref]

[23] Zheng, S., Ristovski, K., Farahat, A., & Gupta, C. (2017, June). Long short-term memory network for remaining useful life estimation. In *2017 IEEE international conference on prognostics and health management (ICPHM)* (pp. 88-95). IEEE. [Crossref]

[24] Wang, J., Wen, G., Yang, S., & Liu, Y. (2018, October). Remaining useful life estimation in prognostics using deep bidirectional LSTM neural network. In *2018 Prognostics and System Health Management Conference (PHM-Chongqing)* (pp. 1037-1042). IEEE. [Crossref]

[25] Liu, C., & Chen, W. (2021). A RUL prediction method of equipments based on MSDCNN-LSTM. *Xibei Gongye Daxue Xuebao/Journal of Northwestern Polytechnical University, 39*(2), 407-413. [Crossref]

[26] Lei, Y., Li, N., Gontarz, S., Lin, J., Radkowski, S., & Dybala, J. (2016). A model-based method for remaining useful life prediction of machinery. *IEEE Transactions on reliability, 65*(3), 1314-1326. [Crossref]

[27] Severson, K. A., Attia, P. M., Jin, N., Perkins, N.,

Jiang, B., Yang, Z., ... & Braatz, R. D. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy, 4*(5), 383-391. [Crossref]

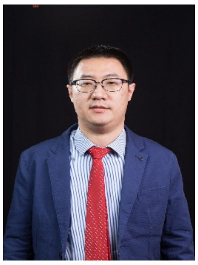
[28] Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.

[29] Weng, C., Cui, Y., Sun, J., & Peng, H. (2013). On-board state of health monitoring of lithium-ion batteries using incremental capacity analysis with support vector regression. *Journal of Power Sources, 235*, 36-44. [Crossref]

[30] Wu, J., Zhang, C., & Chen, Z. (2016). An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks. *Applied energy, 173*, 134-140. [Crossref]

[31] Chang, Y., Fang, H., & Zhang, Y. (2017). A new hybrid method for the prediction of the remaining useful life of a lithium-ion battery. *Applied energy, 206*, 1564-1578. [Crossref]

Xin Zhou received the B.S. degree in Logistics Engineering from the University of Science and Technology Beijing (USTB), China, 2025. (Email: 19923053820@163.com)



Dr. Lechang Yang is a Professor of Reliability Engineering at the University of Science and Technology Beijing (USTB). His research interests include quality and reliability, uncertainty quantification, and PHM. (Email: yanglechang@126.com)