A Novel Progressive Fusion Approach for Remaining Useful Life Prediction Using Temporal-Sensor Data

Gajanan Walunjkar
Department of Information Technology, Army Institute of Technology, Pune, India
gwalunjkar@aitpune.edu.in

Aparna Joshi
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India
aparna.joshi@pccoepune.org (corresponding author)

Rahul Desai
Department of Information Technology, Army Institute of Technology, Pune, India
rahuldesai@aitpune.edu.in

Yuvraj Gholap
Department of Information Technology, Army Institute of Technology, Pune, India
yngholap@aitpune.edu.in

Sachin Korde
Department of Information Technology, Pravara Rural Engineering College, Loni, Ahilyanagar, India
kordesk@pravaraengg.org.in

Received: 6 March 2025 | Revised: 13 May 2025 | Accepted: 1 June 2025

 DOI: https://doi.org/10.48084/etasr.10740

ABSTRACT

Accurate prediction of Remaining Useful Life (RUL) is vital for proactive maintenance and failure prevention in industrial systems. This paper introduces the Spatiotemporal Homogeneous Feature Extractor (STSF) to address limitations in existing RUL prediction methods. STSF employs a flexible, layer-wise progressive feature fusion technique that harmonizes spatial and temporal features, enhancing the model's ability to capture complex degradation patterns. To further improve prediction accuracy, the Feature Space Global Relationship Invariance (FSGRI) training method is used, grounded in supervised contrastive learning. FSGRI maintains consistent relationships between sample features and degradation trajectories during training, simplifying subsequent regression tasks. Evaluations on a C-MAPSS dataset demonstrate that STSF outperforms baseline models across multiple metrics, highlighting its effectiveness in RUL prediction.

Keywords-remaining useful life prediction; spatiotemporal homogeneous feature extractor; feature space; global relationship invariance; feature fusion

I. INTRODUCTION

Monitoring and predicting the Remaining Usable Life (RUL) of machinery and equipment is critical to proactive maintenance and operational efficiency [1]. Traditional methods for RUL prediction often struggle to capture the complex temporal dynamics and cross-variate interactions inherent in multivariate time series data. To address these challenges, this study proposes a Dual-path Mixer (Dual-Mixer) model, which is designed to leverage both temporal and spatial data for improved RUL prediction [2]. The Dual-Mixer model builds on the principles of the TSMixer architecture, which has demonstrated competitive performance in capturing temporal dynamics and cross-variate interactions despite its relatively simplistic design. The core components of TSMixer include residual connections, normalization, temporal projection layers, and multilayer polyphonic filters for feature and time fusion, forming the foundation of the Dual-Mixer. This architecture introduces the Dual-path Mixer Layer

(DML), a key innovation aimed at uniformizing the structure of the features, enabling the fusion and interaction of the adaptable features, and facilitating deep stacking [3]. DML consists of temporal and spatial components, each capable of processing input information of varying sizes. Shared linear mapping layers ensure clear recording of various aspect features, enhancing the model's capacity to extract diverse data. Gate blocks play a crucial role in managing the fusion process by filtering features through interaction links between the temporal and spatial components.

To ensure stability and support deep stacking, residual connections and LayerNorm modules are integrated after the MLP block and feature fusion [4]. This design stabilizes the gradient values during backpropagation, allowing the model to achieve enhanced nonlinear mapping capabilities without compromising the convergence speed. The DML process involves passing input matrices through MLP and Gate blocks in both temporal and spatial portions, followed by feature fusion and normalization procedures. In the feature extraction phase, the Dual-Mixer model employs a linear mapping layer, multiple DML layers, and independent gate blocks to convert raw input samples into a high-dimensional feature space. The final DML layer produces discrete temporal and spatial characteristics, which are then filtered and combined to create an integrated feature. Subsequently, the regression layer processes these comprehensive features to predict RUL [5]. By capturing both temporal and spatial properties, the Dual-Mixer model provides a robust framework for RUL prediction, effectively blending regression and feature extraction. This approach not only improves the predictive accuracy for equipment monitoring, but also offers a versatile and scalable solution for various predictive maintenance tasks.

II. LITERATURE REVIEW

The prediction of RUL has greatly affected several fields in recent years due to the wide adoption of big data techniques, advances in sensor technology, and improved data collection capabilities. The use of deep learning-based data-driven RUL prediction techniques has gained popularity [6]. These techniques take advantage of the abundance of internal data to provide reliable, universal, and approachable solutions. To increase model accuracy and dependability, these techniques have focused mainly on enhancing temporal feature extraction, spatial-temporal fusion feature extraction, and attention networks. Various deep learning techniques have been integrated into the prediction of RUL, leading to recent advances. For more precise temporal feature extraction, a bidirectional Gated Recurrent Unit (GRU) was combined with a temporal attention mechanism in [7]. In [8], state-of-the-art results were achieved using sparse mechanism self-attention modules on a C-MAPSS dataset. In [9], bidirectional GRU networks were combined with a one-dimensional CNN to improve the extraction of spatiotemporal features. These techniques demonstrate the variety of methods used to improve the extraction of temporal and spatial features, which are essential for precise RUL prediction.

For battery RUL prediction, autoencoders, CNN, and LSTM were used in [12] to extract adaptive spatiotemporal features. Furthermore, time series analysis models have become widely used. Transformer architectures, attention mechanisms, and MLP models have been the subject of recent research [10], which has had a great impact on the development of RUL prediction. Slimline MLP-like models, such as Google's TS-Mixer, have demonstrated that less complicated architectures can sometimes outperform intricate transformer networks in terms of performance [14].

TABLE I. LITERATURE SURVEY

Ref.	Focus area/Application	Methods
[9]	RUL prediction (rotating machinery)	Spatial Attention Residual Network
[10]	RUL prediction (Li-ion batteries)	Mode decomposition, Data-driven modeling
[11]	RUL estimation	Deep CNN regression
[12]	RUL prediction (Li-ion batteries)	Auto-CNN-LSTM hybrid model
[13]	Machine fault diagnosis	Wavelet-integrated Residual Network, adversarial training
[14]	RUL prediction framework	Deep learning + stochastic process model
[15]	Fault diagnosis	BLSTM and under-complete autoencoder

III. SPATIAL AND TEMPORAL STREAM FUSION

A. MLP Mixer

This is a complex neural network architecture designed specifically for image processing tasks. Fundamentally, MLP-Mixer uses per-patch linear embeddings, in which each input image patch is linearly transformed to produce an embedding that makes further processing easier. The network consists of mixer layers, which are essential for data integration and feature extraction on both the token and channel dimensions. Two essential parts make up each mixer layer: a token mixing MLP and a channel mixing MLP. Within each token, the Token-Mixing MLP functions by utilizing two fully-connected layers with GELU nonlinearity to efficiently blend information across several channels. In a similar vein, the Channel-Mixing MLP [15] uses two fully connected layers with GELU activation to promote token interaction within the same channel.

TABLE II. MLP-MIXER ARCHITECTURE DETAILS

Component	Details
Input shape	(Batch size, time steps, features) e.g., (B, 128, 16)
Input projection	Dense Layer: 16→64 or 128 neurons
Mixer layers (8–12)	
Token-mixing MLP	Operates across time steps: Input: 128, Hidden: 256, Output: 128
Channel-mixing MLP	Operates across features: Input: 64, Hidden: 512, Output: 64
Normalization	LayerNorm before each MLP
Residual connections	Yes
Output layer	Global average pooling + Dense layer

A two-dimensional input table is produced by each Mixer layer, processing a series of non-overlapping picture patches that are linearly projected to a chosen hidden dimension. Channel-mixing MLPs with parameter tying offer positional invariance, which prevents excessive growth with longer sequences or more hidden dimensions and results in significant memory savings.

B. TS Mixer

The exploration of the capture of time dependencies has shown that linear models are powerful instruments. Multi-Layer Perceptrons (MLPs) stack linear models with nonlinearity to improve their capabilities. Learning is made more efficient by common deep learning approaches such as residual connections and normalization. However, cross-variate information is not taken into account in this original architecture. Interleaving reduces computational complexity and model size while effectively utilizing cross-variational data and temporal relationships. Thus, TSMixer can employ a long lookback window while minimizing parameter growth. For comparison, a simplified variation of just Tmix was used, which just considers time mixing activities. The architecture of TSMixer consists of residual connections, normalization, temporal projection layers, and feature- and time-mixing multilayer polyphonic filters. In real-world circumstances, this design performs competitively despite being relatively simplistic in comparison to recent transformer developments, capturing temporal dynamics and cross-variational interactions adequately. TSMixer has also been expanded to handle additional data, which makes it more flexible and useful for a variety of predictive tasks.

TABLE III. TS-MIXER ARCHITECTURE DETAILS

Component	Details
Input shape	(Batch Size, Time Windows, Features), e.g., (B, 60, 32)
Input projection	Dense Layer: 32→64 or 128 neurons
TS-Mixer layers (4–8)	
Time-mixing MLP	Operates across time: Input: 60, Hidden: 120, Output: 60
Feature-mixing MLP	Operates across features: Input: 64, Hidden: 256, Output: 64
Normalization	LayerNorm
Dropout	0.1-0.3
Residual connections	Yes
Output layer	Dense layer for forecasting or classification

C. STSF Architecture

Designed for multivariate monitoring of time series data, the Dual-path Mixer (Dual-Mixer) model consists of a regression layer and a feature extractor. The regression layer uses the features that were extracted to predict the RUL, while the feature extractor converts input samples from a raw data space to a high-dimensional feature space. A linear mapping layer, several Dual-path Mixer Layers (DML), and independent gate blocks are all included in the feature extraction section. The input samples are first mapped to the feature dimensions of the model by the linear mapping layer. The input for the first layer of DML comes from a single matrix for both the temporal and spatial components, the spatial component being identified using transposition. The final DML layer then produces discrete temporal and spatial characteristics. Before these features are added together to create the final feature, they are further filtered using Gate blocks. By capturing both temporal and spatial properties, this integrated feature makes it easier for the regression layer to process the data later on. The linear mapping layer in the regression section converts the final feature into RUL predictions. All things considered, the Dual-Mixer model efficiently blends regression and feature extraction, providing a thorough method of RUL prediction for equipment monitoring. Three main objectives guide the design of the DML: uniformizing feature structures, enabling adaptable feature fusion and interaction, and making it simple to stack into deep architectures. Figure 1 illustrates its architecture, which consists of temporal and spatial components that are each capable of processing input information of different sizes. The shared linear mapping layers ensure that various aspect features are recorded clearly, improving the model's capacity to extract a variety of data. Gate blocks efficiently manage the fusion process by filtering features before it through interaction links between these components.

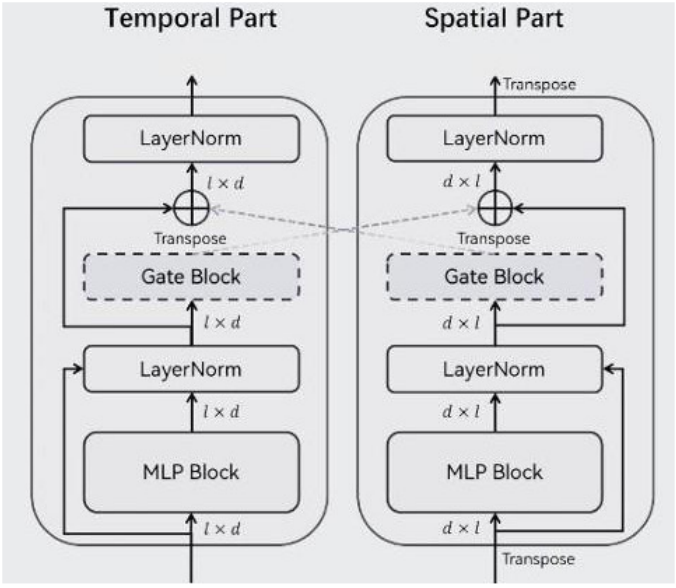

Fig. 1. Spatial and temporal block architecture.

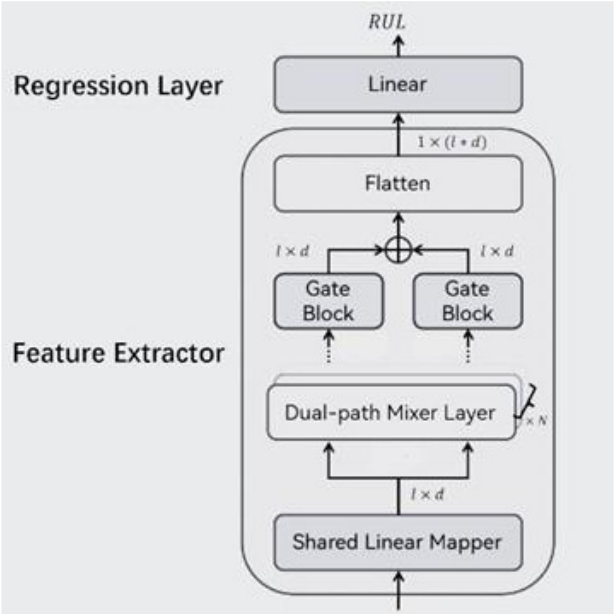

Fig. 2. STSF architecture.

TABLE IV. STFF DETAILS

Component	Details
Input shape	(Batch size, time steps, Sensors) e.g., (B, 100, 20)
Spatial encoder (MLP)	Per time step: Input: 20, Hidden: 64→128
Temporal encoder	LSTM/GRU/Temporal MLP: Hidden Units: 128–256, Layers: 2–3
Fusion layer	Concatenation + Attention/Gating mechanism
Output layer	Dense layer for regression
Dropout	0.1-0.2

IV. EXPERIMENTS

A. Data Preprocessing

A full preprocessing pipeline for the C-MAPSS dataset is essential for training machine learning models, especially for tasks involving RUL prediction. The C-MAPSS dataset, sourced from NASA's Prognostics Data Repository, is used for RUL prediction on aircraft engines [16]. Four key features were selected: Sensor 2 (Fan Inlet Temperature), Sensor 3 (Low-Pressure Compressor Temperature), Sensor 7 (Fan Speed), and Sensor 8 (Engine Pressure Ratio). The dataset was divided into 80% for training and 20% for testing.

Algorithm 1: Preprocessing for RUL prediction:

1. Outlier Removal
   Applied Z-score filtering with a threshold of ±3.0 to remove anomalous data points.
2. Sensor Selection
   Dropped low-variance sensors (e.g., Sensors 1, 5, 10).
   Selected relevant sensors: Sensors 2, 3, 7, 8.
3. Normalization
   Performed min-max scaling to normalize all selected sensor readings to the range [0, 1].
4. RUL calculation
   Computed the RUL for each engine cycle.
   Capped RUL at 125 to handle extremely long life spans and reduce variance.
5. Sliding Window Sampling
   Used a window size of 30 time steps with a stride of 1 to generate overlapping sequences.
   Each window's label is the RUL at the last time step of that window.
6. Train-Test Split
   Performed an 80% training and 20% testing split at the engine unit level to prevent data leakage.
7. Progressive Fusion Preparation:
   Structured the data to feed into the progressive fusion model, combining temporal dependencies with multi-sensor information for robust RUL prediction.

B. Temporal Dependency Handling

TS mixers stand out in predictive maintenance due to their ability to capture temporal dependencies and patterns in sequential data. In the realm of machinery health monitoring, where the progression of conditions over time is paramount, TS mixers excel at discerning changes in sensor readings that lead to potential failure events. By analyzing historical data and identifying subtle shifts or anomalies in the temporal evolution of machinery conditions, TS mixers enable early detection of potential issues, thus facilitating proactive maintenance interventions to prevent costly downtimes and disruptions in industrial operations.

FSGRI is a feature selection mechanism that evaluates the contribution of each input feature to the model's output by computing the gradient-based relevance scores during backpropagation. It identifies the most informative sensor features by analyzing their gradient impact on the RUL prediction, helping to reduce redundant or noisy inputs. In this model, FSGRI was integrated after the QGeoEmbed layer and before the progressive fusion block. This ensures that only the most relevant temporal-sensor features are forwarded to the deeper layers, improving both model interpretability and computational efficiency.

On the other hand, the STSF model adopts a dual MLP architecture, where input features are processed independently in the spatial and temporal domains before being fused for final prediction. The STSF model also offers the option to incorporate an orthogonal regularization loss, promoting feature orthogonality within the model's parameters to improve generalization and mitigate overfitting. Both architectures are designed with scalability and efficiency in mind, leveraging PyTorch's computational graph optimization and GPU acceleration capabilities for accelerated training and inference. Additionally, the code includes functionalities for computing loss functions, handling contrastive learning objectives, and incorporating regularization techniques to enhance model robustness and generalization performance. Overall, the MLP Mixer and STSF models offer powerful tools for predictive maintenance tasks, capable of learning informative representations from time series sensor data and making accurate predictions of machinery health and RUL.

V. RESULTS

The STSF model is distinguished by its lightweight architecture, which provides a simplified substitute for transformer models with more intricate construction. Due to its simplicity, computing resources can be used more effectively, which is especially useful in situations where computational power is scarce or expensive. Despite its simplicity, the STSF model performs remarkably well in identifying complex non-linear relationships in the data. It significantly improves predictive accuracy by focusing on important characteristics and patterns, particularly in regression tasks where comprehending intricate relationships is essential. Additionally, the STSF model reduces the possibility of overfitting, which is a common issue with more intricate designs such as transformers, especially when working with smaller datasets. Due to its simple design, there is less chance of memorizing

noise in the data, which improves generalization to new examples. In addition, its lightweight design speeds up training, which accelerates the model development process. Due to its agility, researchers and practitioners can adjust model parameters and investigate different architectures to maximize performance for particular tasks more quickly. Figures 3-6 present the model performance under different configurations by varying the number of layers (2, 4, 6, etc.) and the embedding dimensions (16, 32, 64, 128). These variations represent different architectural depths and widths tested to find the optimal balance between model complexity and prediction accuracy for RUL estimation. The results show how deeper and wider models capture more complex patterns but may also risk overfitting or higher computational cost.

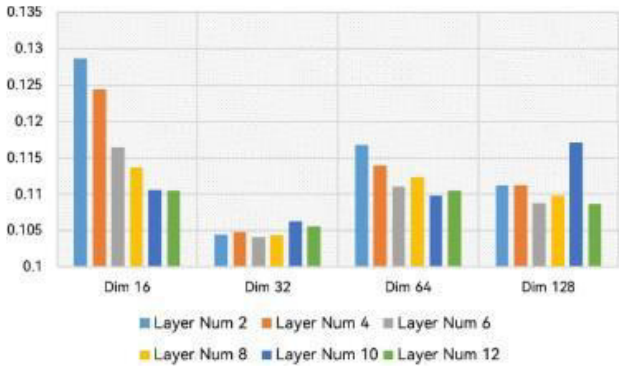

Fig. 3. RMSE obtained by STSF on the FD001 dataset.

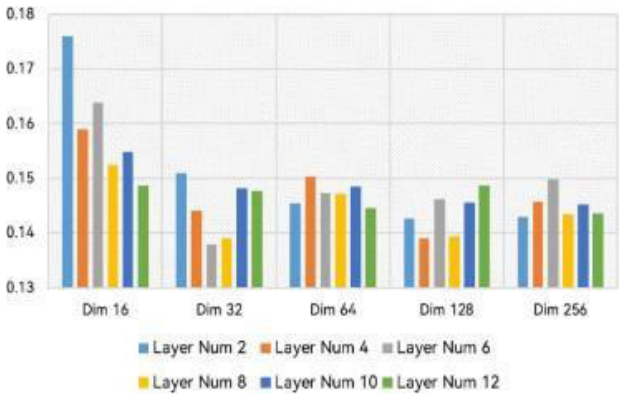

Fig. 4. RMSE obtained by STSF on the FD002 dataset.

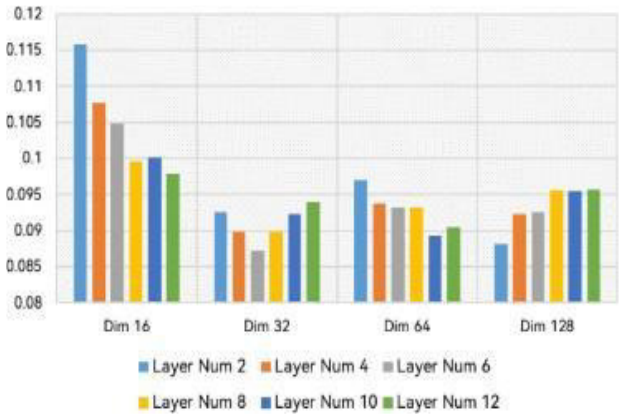

Fig. 5. RMSE obtained by STSF on the FD003 dataset.

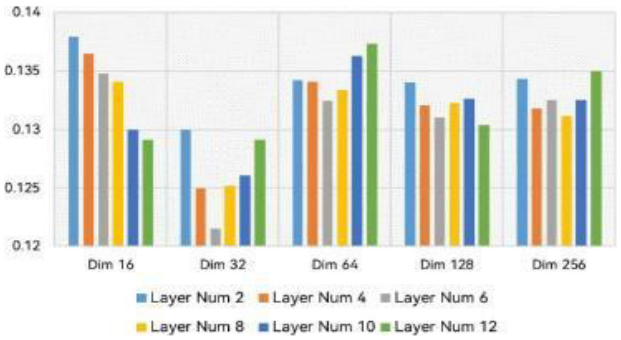

Fig. 6. RMSE obtained by STSF on the FD004 dataset.

Tables V and VI present a comparison of different models for RUL prediction across the FD001-FD004 subsets of the dataset. Models like LSTM, CNN-GRU, DAMCNN, and TS-Mixer were reproduced using the same dataset, 80%-20% split, and standard preprocessing. The proposed STSF model outperformed all baselines, achieving the lowest RMSE and MAPE, confirming its superior accuracy and generalization. The STSF model performs better than other models on a variety of tasks, according to a comparison of models based on RMSE. This shows that for the regression tasks that this study looked at, STSF is the most useful model. Additional examinations, such as the unique characteristics and structure of STSF, can provide a more profound understanding of its exceptional predictive capabilities.

TABLE V. EXPERIMENT RESULTS ON C-MAPSS FD001 AND FD002 DATASETS

Model	**FD001**		**FD002**	
	RMSE	**MAPE**	**RMSE**	**MAPE**
LSTM	0.1279	13.52%	0.1747	20.74%
CNN-GRU	0.1166	12.52%	0.1493	17.30%
DAMCNN	0.1174	12.29%	0.1960	25.45%
BTSAM	0.1062	9.99%	0.1418	18.54%
IMDSSN	0.1100	11.39%	0.1355	16.51%
TS-Mixer	0.1161	12.92%	0.1911	44.50%
STSF (Proposed)	0.1041	10.18%	0.1390	14.98%

TABLE VI. EXPERIMENT RESULTS ON C-MAPSS FD003 AND FD004 DATASETS

Model	**FD003**		**FD004**	
	RMSE	MAPE	RMSE	MAPE
LSTM	0.1116	10.63%	0.1542	16.16%
CNN-GRU	0.0985	9.28%	0.1486	15.68%
DAMCNN	0.0961	8.29%	0.1767	19.10%
BTSAM	0.0925	10.39%	0.1429	12.26%
IMDSSN	0.0959	9.67%	0.1273	13.98%
TS-Mixer	0.1028	9.84%	0.1629	16.56%
STSF (Proposed)	0.0887	7.65%	0.1215	12.36%

VI. CONCLUSION AND FUTURE SCOPE

STSF is a versatile model designed to enhance feature fusion techniques in RUL prediction. By incorporating a spatiotemporal homogeneous feature extractor, the STSF improves predictive performance by effectively capturing complex relationships within the data. In addition, FSGRI aims to smooth the feature space distribution in deep learning-based RUL prediction methods. FSGRI aligns sample feature

relationships with the device degradation process, simplifying the regression task and enhancing prediction accuracy. Through extensive experimentation, the effectiveness of the STSF and FSGRI training method was validated. The results show that STSF outperforms existing methods in various metrics, with the FSGRI training method leading to significant improvements in RMSE and MAPE. Overall, the combination of the STSF and FSGRI training demonstrates superior performance for RUL estimation.

In future work, the STSF model can benefit from the integration of external data sources to enhance prediction accuracy. This exploration could involve incorporating information on environmental conditions or operational parameters, providing a more comprehensive understanding of the system's health. Multitask learning frameworks offer another avenue for exploration, enabling the simultaneous prediction of RUL for multiple components or systems while leveraging shared data for better performance. In addition, developing methods for estimating prediction uncertainty can provide more reliable RUL estimates, crucial for decision-making in maintenance operations. Real-time implementation of the STSF model in industrial settings could facilitate proactive maintenance scheduling and resource optimization, further enhancing operational efficiency. Enhancing the interpretability and explainability of the model, fostering collaboration with domain experts, and conducting comprehensive benchmarking studies are also essential aspects of future research to ensure the relevance and effectiveness of the STSF model in real-world applications. In addition to algorithmic improvements, future work on the STSF model could focus on data-related aspects to enhance predictive performance. This could involve exploring techniques for handling imbalanced datasets and addressing issues related to missing or noisy data, ensuring robustness in real-world scenarios. Furthermore, investigating domain-specific data preprocessing methods tailored to the characteristics of the target system could lead to more effective feature representations and better model performance. Collaborative efforts between data scientists and domain experts can facilitate the development of domain-specific features and insights, contributing to more accurate RUL predictions. Additionally, the scalability and efficiency of the STSF model could be enhanced through parallel computing techniques and optimization strategies, enabling the analysis of larger datasets and faster model training. In general, these directions offer promising avenues for advancing the capabilities of the STSF model and its applicability to practical predictive maintenance scenarios.

REFERENCES

[1] M. Compare, L. Bellani, and E. Zio, "Reliability model of a component equipped with PHM capabilities," *Reliability Engineering & System Safety*, vol. 168, pp. 4–11, Dec. 2017, https://doi.org/10.1016/j.ress.2017.05.024.

[2] S. Wang, Y. Lei, B. Yang, X. Li, Y. Shu, and N. Lu, "A graph neural network-based data cleaning method to prevent intelligent fault diagnosis from data contamination," *Engineering Applications of Artificial Intelligence*, vol. 126, Nov. 2023, Art. no. 107071, https://doi.org/10.1016/j.engappai.2023.107071.

[3] M. Wang, X. Ma, Y. Hu, and Y. Wang, "Gear Fault Diagnosis Based on Variational Modal Decomposition and Wide+Narrow Visual Field Neural Networks," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 4, pp. 3288–3299, Jul. 2022, https://doi.org/10.1109/TASE.2021.3117288.

[4] X. Han, Z. Wang, M. Xie, Y. He, Y. Li, and W. Wang, "Remaining useful life prediction and predictive maintenance strategies for multi-state manufacturing systems considering functional dependence," *Reliability Engineering & System Safety*, vol. 210, Jun. 2021, Art. no. 107560, https://doi.org/10.1016/j.ress.2021.107560.

[5] X. Li *et al.*, "A spectral self-focusing fault diagnosis method for automotive transmissions under gear-shifting conditions," *Mechanical Systems and Signal Processing*, vol. 200, Oct. 2023, Art. no. 110499, https://doi.org/10.1016/j.ymssp.2023.110499.

[6] J. Zhang, Y. Jiang, S. Wu, X. Li, H. Luo, and S. Yin, "Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism," *Reliability Engineering & System Safety*, vol. 221, May 2022, Art. no. 108297, https://doi.org/10.1016/j.ress.2021.108297.

[7] J. Zhang, X. Li, J. Tian, H. Luo, and S. Yin, "An integrated multi-head dual sparse self-attention network for remaining useful life prediction," *Reliability Engineering & System Safety*, vol. 233, May 2023, Art. no. 109096, https://doi.org/10.1016/j.ress.2023.109096.

[8] J. Zhang *et al.*, "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*, vol. 72, pp. 1–12, 2023, https://doi.org/10.1109/TIM.2022.3227956.

[9] W. Xu, Q. Jiang, Y. Shen, Q. Zhu, and F. Xu, "New RUL Prediction Method for Rotating Machinery via Data Feature Distribution and Spatial Attention Residual Network," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–9, 2023, https://doi.org/10.1109/TIM.2023.3246526.

[10] J. Wang, S. Zhang, C. Li, L. Wu, and Y. Wang, "A Data-Driven Method With Mode Decomposition Mechanism for Remaining Useful Life Prediction of Lithium-Ion Batteries," *IEEE Transactions on Power Electronics*, vol. 37, no. 11, pp. 13684–13695, Aug. 2022, https://doi.org/10.1109/TPEL.2022.3183886.

[11] G. S. 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*, 2016, pp. 214–228, https://doi.org/10.1007/978-3-319-32025-0_14.

[12] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen, "A Data-Driven Auto-CNN-LSTM Prediction Model for Lithium-Ion Battery Remaining Useful Life," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3478–3487, Feb. 2021, https://doi.org/10.1109/TII.2020.3008223.

[13] X. Li, Y. Lei, X. Li, and B. Yang, "A Robust Wavelet-integrated Residual Network for Fault Diagnosis of Machines with Adversarial Training," in *2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, Seattle, WA, USA, Jun. 2023, pp. 847–851, https://doi.org/10.1109/aim46323.2023.10196265.

[14] H. Pei, X. Si, T. Li, Z. Zhang, and Y. Lei, "Interactive Prognosis Framework Between Deep Learning and a Stochastic Process Model for Remaining Useful Life Prediction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 12, pp. 18000–18012, Sep. 2024, https://doi.org/10.1109/TNNLS.2023.3310482.

[15] M. Sreenatha and P. B. Mallikarjuna, "A Fault Diagnosis Technique for Wind Turbine Gearbox: An Approach using Optimized BLSTM Neural Network with Undercomplete Autoencoder," *Engineering, Technology & Applied Science Research*, vol. 13, no. 1, pp. 10170–10174, Feb. 2023, https://doi.org/10.48084/etasr.5595.

[16] A. Saxena and K. Goebel, "Turbofan Engine Degradation Simulation." Prognostics Center of Excellence Data Set Repository - NASA, 2008, [Online]. Available: https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/.