Leveraging Time Series Foundation Models Embeddings for Remaining Useful Life Prediction

Ilias Abdouni[1] Alexandre Voisin [2] Christophe Cerisara[3]

[1,2] *Université de Lorraine, CNRS, CRAN, F-54000, NANCY, France*
ilias.abdouni@univ-lorraine.fr
alexandre.voisin@univ-lorraine.fr

[3] *Université de Lorraine, CNRS, LORIA, France*
christophe.cerisara@loria.fr

ABSTRACT

Recent advances in Remaining Useful Life (RUL) prediction rely heavily on task-specific deep learning architectures, such as CNNs, LSTMs, and Transformers. While effective, these data-intensive models frequently struggle to generalize across varying operating conditions. Time Series Foundation Models (TSFMs) offer a promising zero-shot alternative to this training-from-scratch approach, yet directly applying their forecasting-optimized representations to prognostics often fails to capture the physical constraints of equipment degradation. To resolve this task-objective mismatch, we propose a domain-agnostic adapter architecture that applies the Wide & Deep learning paradigm to repurpose the frozen Chronos-2 foundation model for continuous RUL regression. Our methodology explicitly bridges the domain gap by extracting and flattening the model's abstract multivariate embeddings (Deep), and fusing them with raw, normalized physical measurements (Wide). Experiments on the full C-MAPSS benchmark demonstrate that this approach achieves state-of-the-art performance, reaching an average RMSE of 10.32. By outperforming recent specialized architectures without requiring backbone fine-tuning, this work proves that lightweight adaptation of generalized temporal representations offers a scalable, robust alternative to traditional prognostic modeling.

1. INTRODUCTION

In recent years, Data-Driven approaches based on Deep Learning (DL) have established themselves as the state-of-the-art for Remaining Useful Life (RUL) prognostics. Architectures ranging from Convolutional Neural Networks (CNNs) to Long Short-Term Memory (LSTM) networks and Transformers have demonstrated remarkable accuracy on standard benchmarks (Zheng, Ristovski, Farahat, & Gupta, 2017; L. Liu, Song, & Zhou, 2022). However, these dominant approaches frequently rely on a data-intensive, training-from-scratch paradigm. While Transfer Learning (TL) techniques are often used to mitigate this, they traditionally depend on strictly aligning a labeled source domain with a highly similar target domain to avoid negative transfer (Zhuang et al., 2020). Consequently, standard prognostic architectures remain structurally tied to their training distributions, making them highly prone to overfitting when labeled run-to-failure data is scarce. Furthermore, these models often struggle to maintain accuracy when subjected to noise and varying operating conditions, frequently requiring task-specific redesigns to extract a cohesive degradation signature.

Simultaneously, the field of general time series analysis and prediction has been transformed by the emergence of Foundation Models (TSFMs) such as Chronos (Ansari et al., 2024a), Moment (Goswami et al., 2024), and TimeGPT (Garza, Challu, & Mergenthaler-Canseco, 2023). Pre-trained on large and diverse collections of temporal data, these models learn transferable temporal representations and exhibit strong zero-shot and few-shot forecasting capabilities. In principle, their internal representations provide a promising basis for prognostics, reducing reliance on large labeled task-specific datasets.

Yet, directly applying TSFMs to RUL prediction remains challenging. While models like Chronos-2 support multivariate reasoning via group attention, their internal representations are optimized for forecasting future values rather than estimating a unified physical degradation curve. This creates a task-objective mismatch. Straightforward application frequently leads to suboptimal performance because generalized forecasting attention maps cannot easily distill the compact, degradation-specific signature required for prognostics. However, by extracting these high-dimensional multivariate embeddings, we can design a specialized downstream adapter to

adapt these features for regression-specific latent spaces.

In this paper, we propose a novel application that repurposes the Chronos-2 foundation model for RUL estimation. Our specific contributions are as follows:

- **Zero-Shot Feature Extraction:** We demonstrate that the frozen Chronos-2 encoder yields robust, degradation-relevant features without requiring task-specific fine-tuning.
- **Compress-and-Flatten Adaptation:** We propose a lightweight adapter that maps high-dimensional forecasting embeddings into a regression-specific, multivariate latent space.
- **Wide & Deep Fusion:** We introduce a dual-path architecture that effectively grounds abstract temporal embeddings with raw physical measurements and explicit regime contexts.
- **State-of-the-Art Performance:** Evaluated on the full C-MAPSS benchmark, our frozen-backbone method achieves an average RMSE of 10.32, outperforming recent specialized models.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 details our proposed architecture. Section 4 presents the experimental setup and results. Finally, Section 5 concludes the paper.

2. RELATED WORK

2.1. Task-Specific Deep Learning for RUL

Data-driven RUL estimation has evolved from standard sequence modeling to complex representation learning. Early standards like Deep Convolutional Neural Networks (DCNN) and Bidirectional LSTMs (BiLSTM) established baselines for capturing spatial and temporal dependencies (Zheng, Ristovski, et al., 2017; X. Li, Ding, & Sun, 2018). To improve feature extraction without manual selection, architectures like Multi-Scale DCNNs (H. Li, Zhao, Zhang, & Zio, 2020) and end-to-end MLP-LSTM-MLP networks (Chaoub et al., 2022) were introduced. Subsequently, dual-attention models (DA-LSTM (Shi et al., 2024), DA-Transformer (L. Liu et al., 2022)) emerged to dynamically weight critical sensors and time steps.

Recently, the field has adopted long-range sequence models, such as Structured State Space (S4) (Gu, Goel, & Ré, 2022) and hybrid Mamba-Attention (MambaAtt) (Han, Kwon, & Yoon, 2025), to efficiently process extended degradation histories. Concurrently, explicitly constrained networks like GOLCRL (Y. Li et al., 2025) and DMDCN (Y. Wang, Li, Liu, & Cong, 2026) have been introduced to enforce physical and geometric consistency within the latent space. Despite high precision, these models are typically trained in a data-intensive manner, often from scratch, limiting their ability to generalize across diverse operating conditions without task-specific fine-tuning.

2.2. Time Series Foundation Models

Inspired by the success of Large Language Models, sequence modeling has increasingly shifted toward Time Series Foundation Models (TSFMs) (Kim, Kim, Kim, Lee, & Yoon, 2025). This transition is largely driven by the limitations of traditional deep learning architectures, which rely heavily on extensive, task-specific datasets (Miller et al., 2024). In contrast, TSFMs leverage massive multi-domain pre-training to learn universal representations, enabling robust zero-shot forecasting across unseen domains. As these architectures have increased, recent taxonomies (Ye et al., 2025) have begun categorizing native TSFMs based on how they ingest continuous data: they either group time steps into patches to preserve local context, or they discretize values through quantization-based tokenization (Kottapalli et al., 2025). The Chronos architecture (Ansari et al., 2024b) is an example of this quantization principle, reframing forecasting as a probabilistic language modeling task. Building upon this foundation, Chronos-2 advances the framework by integrating a multivariate group attention mechanism. Yet, despite these zero-shot capabilities, a structural issue remains. Because these native TSFMs are fundamentally optimized to predict probability distributions over a discrete vocabulary of future tokens (Kottapalli et al., 2025), a task-objective mismatch arises when applying them to Remaining Useful Life estimation—a task that demands deterministic, continuous regression.

2.3. Foundation Models in Predictive Maintenance

Applying TSFMs to industrial diagnostics offers novel adaptability but faces challenges like data heterogeneity and deployment latency (Ayyat, Osman, & Nadeem, 2025). Current approaches to overcome this include domain-specific pre-training (e.g., GPT4Battery (Feng, Hu, & Zhang, 2024), RmGPT (Y. Wang et al., 2025)), unified multi-task frameworks (Yu, Zhu, Wang, & Tsung, 2025), and LLM adaptations via prompt-tuning (X. Wang & Wang, 2024; J. Liu et al., 2024).

However, directly applying general-purpose models like Moment or Chronos-2 to RUL estimation remains non-trivial. While TSFM embeddings show promise in few-shot scenarios (Dintén & Zorrilla, 2025), naive applications frequently underperform optimized specialized models. For instance, Benzine et al. (Benzine et al., 2025) demonstrated that the original Chronos-2 model struggled significantly on the C-MAPSS dataset. Traditionally, deep learning methods address multivariate cross-sensor correlations through computationally intensive mechanisms, such as Graph Convolutional Networks (GCNs) or spatial-temporal cross-attention layers. To provide a more lightweight alternative for TSFMs, our work introduces a "Compress-and-Flatten" adapter that explicitly aggregates

forecasting tokens into a cohesive multivariate representation for efficient RUL regression.

3. METHODOLOGY

We propose a hybrid architecture that combines the zero-shot representational power of Time Series Foundation Models (TSFMs) with domain-specific physical context. As illustrated in Figure 1, our architecture processes data through two parallel streams: a "Deep Branch" that extracts temporal semantics via a frozen Chronos-2 encoder, and a "Wide Branch" that preserves raw physical measurements and operating conditions.

3.1. Regime-Aware Preprocessing

Complex degradation processes involve varying operating regimes that shift sensor distributions. To inform the model of the physical regime, we apply K-Means clustering (MacQueen, 1967) on the three operational settings $\mathbf{s}_t \in R^3$ (which correspond to altitude, Mach number, and throttle resolver angle). K-Means was selected as a standard, lightweight method for regime identification, discretizing the continuous operational space into K discrete operating conditions. The continuous settings $\mathbf{s}_t$ map to a cluster index k, which is then converted into a one-hot vector $\mathbf{c}_t \in \{0,1\}^K$.

To ensure the "Wide Branch" receives statistically stable inputs, we employ a regime-specific normalization. For each sensor j and cluster k, we compute the mean $\mu_{j,k}$ and standard deviation $\sigma_{j,k}$ from training samples in that regime. The raw sensor input $x_{t,j}$ is standardized as:

$$\tilde{x}_{t,j} = \frac{x_{t,j} - \mu_{j,k}}{\sigma_{j,k}} \tag{1}$$

3.2. Deep Branch: Chronos-2 Feature Extraction

We utilize the pre-trained Chronos-2 (T5-Encoder) as a frozen feature extractor. Taking the standardized multivariate sensor sequences $\mathbf{X}_t \in R^{T\times N}$ from the preprocessing step, the input is processed jointly utilizing Chronos-2's native group attention mechanism. The encoder $\mathcal{E}$ produces a high-dimensional embedding for the last time-step of each sensor:

$$\mathbf{E} = \mathcal{E}(\mathbf{X}_t) \in R^{N\times D_{model}} \tag{2}$$

where $D_{model} = 768$. Thanks to the group attention layers, $\mathbf{E}$ captures rich multivariate temporal dynamics. However, because these representations are fundamentally optimized for general forecasting rather than physical degradation tracking, we extract these native embeddings and pass them to our downstream adapter to specialize them for prognostics.

3.3. Deep Branch: The Compress-and-Flatten Adapter

To bridge the gap between general forecasting representations and physical degradation tracking, we introduce a lightweight adapter module consisting of compression and flattening steps. We first filter forecasting-specific noise and distill the representations by projecting each sensor's 768-dimensional embedding into a compact latent space of size d_{comp} (where $d_{comp} \ll D_{model}$):

$$\mathbf{z}^{(i)} = \sigma(\mathbf{W}_{comp}\mathbf{E}^{(i)} + \mathbf{b}_{comp}) \in R^{d_{comp}} \tag{3}$$

where σ is the Gaussian Error Linear Unit (GELU) activation. GELU was chosen over standard ReLU because its non-zero gradient for negative values prevents "dead neurons" and provides a smoother mapping within the compressed latent space.

Unlike standard approaches that aggregate embeddings via mean or max pooling, we perform a flattening operation to preserve the structural position of each sensor's feature representation. The compressed vectors $\mathbf{z}^{(i)}$ from all N sensors are concatenated:

$$\mathbf{v}_{deep} = \text{Flatten}([\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(N)}]) \in R^{N\cdot d_{comp}} \tag{4}$$

This transforms the 'channel' dimension into a 'feature' dimension, enabling subsequent MLP layers to explicitly map the pre-trained temporal features into the specific physical degradation signatures required for RUL estimation.

3.4. Wide & Deep Fusion

The final regression fuses abstract temporal features with precise physical measurements. We construct a fused vector $\mathbf{v}_{fusion}$ by concatenating the deep temporal embeddings $\mathbf{v}_{deep}$, the regime-normalized raw sensor readings $\tilde{\mathbf{x}}_t$, and the operating condition vector $\mathbf{c}_t$:

$$\mathbf{v}_{fusion} = [\mathbf{v}_{deep} \parallel \tilde{\mathbf{x}}_t \parallel \mathbf{c}_t] \tag{5}$$

Finally, $\mathbf{v}_{fusion}$ is passed through a Multi-Layer Perceptron (MLP) regressor to predict the final RUL estimate $\hat{y}_t$.

4. EXPERIMENTS

To evaluate our architecture, we conducted experiments on the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) benchmark (FD001–FD004) (Saxena, Goebel, Simon, & Eholzer, 2008). Testing on all sub-datasets assesses the model's performance under both constant and varying operating conditions.

4.1. Dataset Description

The C-MAPSS benchmark consists of four sub-datasets (FD001–FD004) containing run to failure trajectories of turbofan engines under varying operating conditions and fault modes. As detailed in Table 1, the datasets range from single-regime scenarios (FD001) to highly complex, multi-regime environments (FD004).

RUL Labeling: During the early operational cycles of a sys-

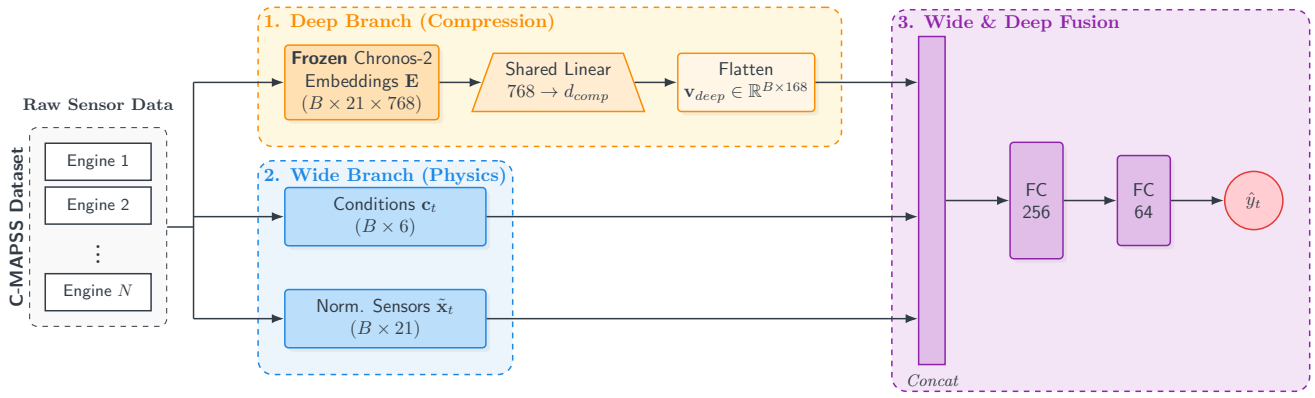

Figure 1. The proposed Wide & Deep Architecture. The Deep Branch (top) extracts and compresses temporal embeddings from Chronos-2, while the Wide Branch (bottom) injects explicit physical context via raw sensors and operating condition clusters.

Table 1. Statistical summary of the C-MAPSS datasets.

Dataset	FD001	FD002	FD003	FD004
Train / Test Trajectories	100 / 100	260 / 259	100 / 100	249 / 249
Operating Conditions	1	6	1	6
Fault Modes	1	1	2	2
Trajectory Length Distribution (Max / Mean / Min)				
Lengths	360/206/128	378/206/128	525/247/145	543/245/128

tem, physical degradation is generally imperceptible. To reflect this, we employ a piece-wise linear target RUL y_t at time step t, capped at a maximum value of $R_{max} = 125$. Originally introduced by Heimes (Heimes, 2008) and subsequently standardized at 125 cycles by Zheng et al. (Zheng, Ruelens, Valkenborg, & Lemmens, 2017), this limit is widely adopted in C-MAPSS literature. It prevents the network from over-penalizing predictions during the early-life phase, where the physical degradation remains unobservable:

$$y_t = \min(R_{max}, T - t) \tag{6}$$

where T is the total cycle life.

Operating Condition Clustering: We apply K-Means clustering ($K = 6$) to the control settings for complex datasets. The value $K = 6$ was chosen as it matches the known six operational conditions present in the FD002 and FD004 datasets. These clusters serve two purposes:

1. **Deep Branch:** Guide Z-score normalization to ensure stable inputs.
2. **Wide Branch:** Serve as explicit one-hot encoded inputs ($\mathbf{c}_t$) for the regressor.

Feature Engineering: Unlike prior literature, which often relies on handcrafted features, fixed-length sliding windows, or zero-padding strategies, our method utilizes all 21 available sensors and processes the entire trajectory history natively without manual feature extraction.

4.2. Implementation Details

Training Strategy: We split the official training set into an 80% training and 20% validation subset, strictly partitioned by Engine ID to prevent data leakage. The Deep Branch utilizes the Chronos-2 Base encoder ($d_{model} = 768$) with frozen weights.

Hyperparameters: Through a grid search ($d_{comp} \in \{2, 4, 8, 16, 128, 256\}$), we set the compression dimension to $d_{comp} = 8$. The fused vector (168 features + wide features) is processed by an MLP (256, 64 units) with GELU and Dropout (0.2). The network is trained using Mean Squared Error (MSE) loss. The model is optimized using AdamW (learning rate $1e-4$, weight decay $1e-3$) and batch size 1024 for up to 150 epochs, with early stopping (patience=15) based on validation RMSE.

4.3. Evaluation Metrics

We employ the official C-MAPSS metrics. Root Mean Squared Error (RMSE) penalizes large errors equally:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{7}$$

The asymmetric Scoring Function (S) heavily penalizes late predictions (where $d_i = \hat{y}_i - y_i$):

$$S = \sum_{i=1}^{N} s_i, \quad s_i = \begin{cases} e^{-\frac{d_i}{13}} - 1 & \text{for } d_i < 0 \\ e^{\frac{d_i}{10}} - 1 & \text{for } d_i \geq 0 \end{cases} \tag{8}$$

4.4. Experimental Results

4.4.1. Model Performance and Stability

To evaluate the robustness of the proposed framework, we conducted repeated experiments ($N = 10$ runs) with different

random seeds. Table 2 summarizes the Mean and Standard Deviation (Std) for both metrics. The low standard deviations (e.g., ± 0.17 on FD004 RMSE) indicate high stability, attributed to the frozen foundation model encoder providing a consistent, high-quality feature space regardless of regression head initialization.

Table 2. Performance Analysis (Mean $\pm$ Std over 10 runs) of the proposed framework.

Dataset	**RMSE** (Mean $\pm$ Std)	**Score** (Mean $\pm$ Std)
FD001	9.97 ± 0.34	187.4 ± 15.4
FD002	11.16 ± 0.21	693.1 ± 74.6
FD003	9.31 ± 0.31	175.9 ± 9.6
FD004	10.83 ± 0.17	589.4 ± 18.8

4.4.2. Comparative Analysis

To evaluate efficacy, we benchmark the *Wide & Deep Chronos-2 Adapter* against three generations of prognostic architectures: Classical Deep Learning (e.g., BiLSTM, MLP-LSTM-MLP), Advanced Sequence and State-Space Models (e.g., DA-Transformer, MambaAtt, DMDCN), and Foundation Model adaptations (e.g., Chronos-2, Moment).

Table 3 presents the performance metrics. Our method yields the lowest RMSE across all datasets.

The experimental results illustrate a significant generational leap in prognostic performance. While classical models such as MS-DCNN struggle under the multiple operating conditions of FD002 and FD004 (average RMSE 16.17), and advanced architectures like DMDCN offer incremental improvements (average RMSE 11.74), our Wide & Deep adaptation achieves a noticeable amelioration in performance. By reducing the average RMSE to 10.32, our model achieves a 12.1% improvement over the current state-of-the-art. Notably, on the highly complex multi-regime FD004 dataset, our method yields an RMSE of 10.83, outperforming the closest baseline (GOLCRL at 12.53) by 13.5%.

This performance suggests that the semantic richness of a Foundation Model, pre-trained on billions of tokens, can compensate for the specialized architectural advantages of scratch-trained hybrid models like MambaAtt. Furthermore, the contrast between our results and the limitations of naive TSFM applications,specifically reducing the 65.11 RMSE reported by Benzine et al. on FD001 down to 9.97 (an 84.6% error reduction), justifies the necessity of our specific adaptation strategy. By fusing frozen temporal embeddings with raw sensor physics, we successfully bridge the domain gap that previously hindered the application of native foundation models to multivariate industrial regression.

4.4.3. Ablation Study: The Impact of the Wide Branch

To demonstrate the value of our Wide & Deep architecture, we performed an ablation study (Table 4) comparing the foundation model embeddings alone ("Only Deep") versus fusing them with raw physical data and conditions ("Full Fusion").

The "Only Deep" configuration serves as the native multivariate foundation model baseline, relying exclusively on the joint multivariate embeddings extracted via Chronos-2's group attention, processed by the adapter's prediction head. This baseline confirms the robustness of Chronos-2's forecasting representations. However, the "Full Fusion" model demonstrates consistent improvements across all datasets. This proves that while Chronos-2 captures the semantic degradation trend, the "Wide" branch is strictly necessary to restore the fine-grained numerical precision and operating context required for accurate RUL regression.

4.4.4. Temporal Analysis: Run to Failure Trajectory

To assess temporal behavior, Figure 2 compares the predicted RUL trajectory with the ground truth for an unseen engine in the most challenging multi-regime dataset (FD004).

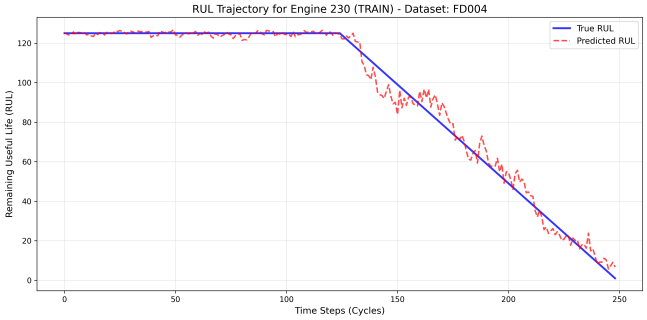

Figure 2. Run to failure prediction on an unseen validation engine (FD004, Unit 230). The prediction (red dashed line) remains stable and closely follows the ground truth (blue solid line) despite the high variance of operating conditions.

The prediction remains remarkably smooth despite frequent changes in operating conditions. In most architectures, sudden shifts in raw sensor measurements caused by operating conditions changes often lead to unstable, irregular RUL predictions. This smooth trajectory confirms that the Wide & Deep architecture successfully utilizes the explicit condition clusters to isolate the underlying degradation signal from operational noise.

5. CONCLUSION AND FUTURE WORK

In this paper, we presented a novel architecture that repurposes Chronos-2 as a robust feature extractor for Remaining Useful Life estimation. By introducing a "Compress-and-Flatten" adapter and a Wide & Deep fusion strategy, we bridged the gap between the model's generalized forecasting representations

Table 3. Comprehensive performance comparison.

		FD001		FD002		FD003		FD004		Average	
Category	**Method**	**RMSE**	**Score**	**RMSE**	**Score**	**RMSE**	**Score**	**RMSE**	**Score**	**RMSE**	**Score**
4**Classic DL*	BiLSTM (Zheng, Ristovski, et al., 2017)	13.65	295	23.18	4130	13.74	317	24.68	5430	18.81	2543
	DCNN (X. Li et al., 2018)	12.61	274	22.36	10412	12.64	284	23.31	12466	17.73	5859
	MS-DCNN (H. Li et al., 2020)	11.44	196	19.35	3747	11.67	242	22.22	4844	16.17	2257
	MLP-LSTM-MLP (Chaoub et al., 2022)	13.26	285	12.49	**571**	13.11	352	13.97	1252	13.21	615
6**Advanced SOTA*	S4 (Gu et al., 2022)	14.81	271	14.44	933	13.75	261	19.50	28513	15.62	7494
	DA-LSTM (Shi et al., 2024)	12.62	263	13.22	842	13.34	360	16.25	1372	13.86	709
	DA-Transformer (L. Liu et al., 2022)	12.25	198	17.08	1575	13.39	290	19.86	1741	15.64	951
	MambaAtt (Han et al., 2025)	11.46	201	13.20	717	11.56	179	13.25	1114	12.36	552
	LECformer (Zhang et al., 2025)	11.31	204	12.97	705	11.05	195	16.93	1309	13.07	603
	GOLCRL (Y. Li et al., 2025)	11.91	218	12.80	804	11.21	237	12.53	745	12.11	501
	DMDCN (Y. Wang et al., 2026)	10.89	153	12.06	390	11.21	237	12.81	364	11.74	286
3**Foundation Models*	Benzine et al. (Benzine et al., 2025)	65.11	-	-	-	-	-	-	-	-	-
	Dintén et al. (Dintén & Zorrilla, 2025)	15.17	1655	19.79	2641	13.18	2061	16.37	6032	16.13	3097
	Ours (Wide & Deep)	**9.97**	**187**	**11.16**	693	**9.31**	**176**	**10.83**	**589**	**10.32**	**411**

Table 4. Ablation of the Wide Branch (RMSE Mean $\pm$ Std).

Model	**FD001**	**FD002**	**FD003**	**FD004**
Only Deep	10.68 ± 0.34	11.65 ± 0.23	10.16 ± 0.19	11.19 ± 0.27
Full Fusion	$\mathbf{9.97 \pm 0.34}$	$\mathbf{11.16 \pm 0.21}$	$\mathbf{9.31 \pm 0.31}$	$\mathbf{10.83 \pm 0.17}$

and the physics-constrained nature of prognostics.

Our experiments on the full C-MAPSS benchmark demonstrate state-of-the-art performance (Average RMSE 10.32 over all subdatasets). These results suggest that Foundation Models, adapted via task-specific adapters, offer a viable and scalable alternative to training prognostic models from scratch. This represents a promising first step toward translating the foundational model success seen in other domains into predictive maintenance.

Future work will focus on validating this architecture on datasets characterized by higher noise levels and irregular sampling rates, such as N-CMAPSS. Additionally, because the proposed architecture relies on a heavy frozen backbone combined with a lightweight trainable adapter, it presents distinct advantages for privacy-preserving distributed training in Federated Learning scenarios. Finally, as the edge deployment of heavy TSFMs introduces critical latency bottlenecks, future research will explore model distillation techniques and explainability tools-such as sensor importance mapping-to ensure real-time deployment and build trust in our proposed solution.

DATA AVAILABILITY

The C-MAPSS benchmark is publicly available. The code and pre-trained adapter weights to reproduce these experiments will be open-sourced upon publication.

ACKNOWLEDGMENTS

This research work has been conducted within the framework of the MODAPTO project (MODULAR MANUFACTURING AND DISTRIBUTED CONTROL VIA INTEROPERABLE DIGITAL TWINS), which has received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101091996.

Generative AI tools were used to assist with language refinement and formatting of this manuscript.

REFERENCES

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., ... others (2024a). Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815*.

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., ... others (2024b). Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815*.

Ayyat, M., Osman, M., & Nadeem, T. (2025). Opportunities and challenges of foundation models in industrial manufacturing. *IEEE Access*.

Benzine, F. K., Beldjoudi, S., Benzine, A., Zouaghi, I., Haouari, F., & Seridi, H. (2025). From lstm to chronos: evaluating deep learning models for predictive maintenance. In *2025 intelligent methods, systems, and applications(imsa)* (pp. 12–17).

Chaoub, A., et al. (2022). An end-to-end mlp-lstm-mlp model for remaining useful life estimation. *Prognostics and Health Management Conference*.

Dintén, R., & Zorrilla, M. (2025). Using time series foundation models for few-shot remaining useful life prediction of aircraft engines. *Computer Modeling in Engineering & Sciences (CMES)*, *144*(1).

Feng, Y., Hu, G., & Zhang, Z. (2024). Gpt4battery: An llm-driven framework for adaptive state of health estimation of raw li-ion batteries. *arXiv preprint arXiv:2402.00068*.

Garza, A., Challu, C., & Mergenthaler-Canseco, M. (2023). Timegpt-1. *arXiv preprint arXiv:2310.03589*.

Goswami, M., Szafer, K., Choudhry, A., Cai, Y., Li, S., & Dubrawski, A. (2024). Moment: A family of

open time-series foundation models. *arXiv preprint arXiv:2402.03885*.

Gu, A., Goel, K., & Ré, C. (2022). Efficiently modeling long sequences with structured state spaces. In *International conference on learning representations (iclr)*.

Han, J., Kwon, Y., & Yoon, H. (2025). Mamba-attention: A self-supervised framework for efficient remaining useful life prediction. *Reliability Engineering & System Safety*, 111492.

Heimes, F. O. (2008). Recurrent neural networks for remaining useful life estimation. In *2008 international conference on prognostics and health management* (pp. 1–6).

Kim, J., Kim, H., Kim, H., Lee, D., & Yoon, S. (2025). A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges. *Artificial Intelligence Review*, *58*(7), 1–95.

Kottapalli, S. R. K., Hubli, K., Chandrashekhara, S., Jain, G., Hubli, S., Botla, G., & Doddaiah, R. (2025). Foundation models for time series: A survey. *arXiv preprint arXiv:2504.04011*.

Li, H., Zhao, W., Zhang, Y., & Zio, E. (2020). Remaining useful life prediction using multi-scale deep convolutional neural network. *Applied Soft Computing*, *89*, 106113.

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.

Li, Y., Hu, C., Zhou, Z., Sun, C., Peng, J., & Yan, R. (2025). Learning globally ordered and locally consistent degradation representations for remaining useful life prediction. *Advanced Engineering Informatics*, *68*, 103692.

Liu, J., et al. (2024). Large language models as general pattern machines for industrial prognostics. *Reliability Engineering & System Safety*.

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.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297). University of California Press.

Miller, J. A., Aldosari, M., Saeed, F., Barna, N. H., Rana, S., Arpinar, I. B., & Liu, N. (2024). A survey of deep learning and foundation models for time series forecasting. *arXiv preprint arXiv:2401.13912*.

Saxena, A., Goebel, K., Simon, D., & Eholzer, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. *International Conference on Prognostics and Health Management*.

Shi, J., Zhong, J., Zhang, Y., Xiao, B., Xiao, L., & Zheng, Y. (2024). A dual attention lstm lightweight model based on exponential smoothing for remaining useful life prediction. *Reliability Engineering & System Safety*, *243*, 109821.

Wang, X., & Wang, Y. (2024). Pulse: Prompt-based latent space encoder for time series. *arXiv preprint*.

Wang, Y., Li, M., Liu, S., & Cong, H. (2026). Degradation manifold dynamic consistency network for probabilistic remaining useful life prediction. *Nondestructive Testing and Evaluation*, 1–27.

Wang, Y., Yu, Y., Sun, K., Lei, P., Zhang, Y., Zio, E., … Li, Y. (2025). Rmgpt: A foundation model with generative pre-trained transformer for fault diagnosis and prognosis in rotating machinery. *IEEE Internet of Things Journal*.

Ye, J., Yu, Y., Zhang, W., Wang, L., Li, J., & Tsung, F. (2025). *Empowering time series analysis with foundation models: A comprehensive survey.* Retrieved from `https://arxiv.org/abs/2405.02358`

Yu, Y., Zhu, F., Wang, D., & Tsung, F. (2025). Empowering phm applications with time series foundation models: A unified multi-task learning approach. In *2025 ieee 21st international conference on automation science and engineering (case)* (pp. 3300–3305).

Zhang, Z., Song, W., Wu, Q., Sun, W., Li, Q., & Jia, L. (2025). A novel local enhanced channel self-attention based on transformer for industrial remaining useful life prediction. *Engineering Applications of Artificial Intelligence*, *141*, 109815.

Zheng, S., Ristovski, K., Farahat, A., & Gupta, C. (2017). Long short-term memory network for remaining useful life estimation. In *2017 ieee international conference on prognostics and health management (icphm)* (pp. 88–95).

Zheng, S., Ruelens, K., Valkenborg, D., & Lemmens, B. (2017). Long short-term memory network for remaining useful life estimation. In *2017 ieee international conference on prognostics and health management (icphm)* (pp. 88–95).

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., … He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, *109*(1), 43–76.