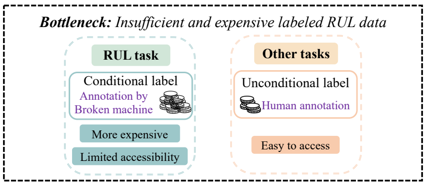
Remaining Useful Life Prediction via Frequency Emphasizing Mix-Up and Masked Reconstruction

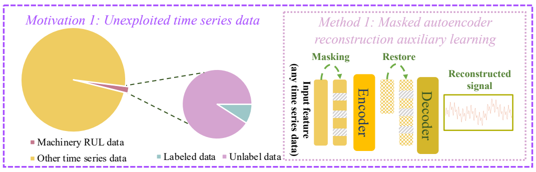
IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE, VOL. 5, NO. 9, SEPTEMBER 2024

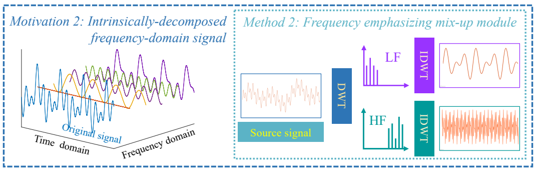
Haoren Guo[1], Haiyue Zhu[2], *Member, IEEE,* Jiahui Wang[1], Vadakkepat Prahlad[1], *Senior Member, IEEE,* Weng Khuen Ho[1], Clarence W. de Silva[3], *Life Fellow, IEEE,* and Tong Heng Lee[1]

[1]College of Design and Engineering, National University of Singapore, [2]Singapore Institute of Manufacturing Technology (SIMTech), Agency for Science, Technology and Research(A*STAR), [3]Department of Mechanical Engineering, University of British Columbia

Motivations







Contributions

- Designed **an** **LSTM temporal projection** layer to effectively capture **long-term sequential information** in time-series datasets.
- Proposed a learning framework for RUL prediction that includes the **FEMM module**, leveraging **decomposed frequency-domain information** to enhance feature extraction.
- Introduced the **MARAL module**, incorporating self-supervised learning (SSL) to utilize **unlabeled data from unrestricted domains**, addressing data scarcity and improving feature extraction and generalization in RUL prediction.

Problem Formulation

Fully-supervised Learning

$$\hat{y} = \mathcal{F}_{\Theta}(X|\mathcal{D}^l)$$

$$\mathcal{D}^l = \{(X_i, y_i)|i = 1, \dots, N\}$$

Semi-supervised Learning (SSL)

$$\hat{y} = \mathcal{F}_{\Theta}(X|\mathcal{D}^l \cup \mathcal{D}^{ul})$$

$$\mathcal{D}^{ul} = \{(X_i)|i = 1, \dots, M\}$$

N: Number of RUL sample pairs M: Number of unlabel samples
$\mathcal{F}_{\Theta}(\cdot)$: Prediction function

Methodology

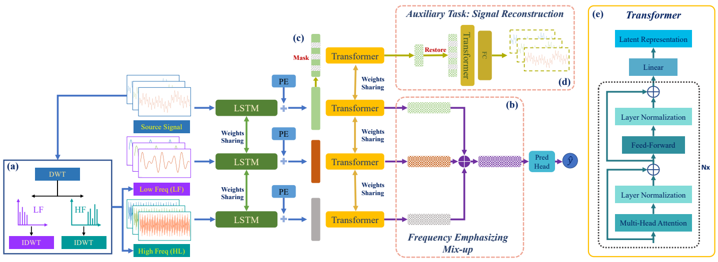


Figure 1. RUL Prediction Framework

Frequency Emphasizing Mix-up Module (FEMM)

Part (a)

Decomposed the source signal into high and low frequency by Discrete Wavelet Transform (DWT).

$$\overline{\mathbf{X}}, \underline{\mathbf{X}} = D_{DWT}(\mathbf{X}) \;\; \mathbf{X}, \overline{\mathbf{X}}, \underline{\mathbf{X}} \in \mathbb{R}^{P\times J}$$

LSTM Temporal Projection Layer

$$\mathbf{H}/\overline{\mathbf{H}}/\underline{\mathbf{H}} = T(\mathbf{X}/\overline{\mathbf{X}}/\underline{\mathbf{X}})$$

Positional Encoding

$$PE_{(z,2d)} = \sin\left(z/10000^{2d/D}\right),$$
$$PE_{(z,2d+1)} = \cos\left(z/10000^{2d/D}\right)$$

Transformer Encoder

$$\mathbf{R}/\overline{\mathbf{R}}/\underline{\mathbf{R}} = TE(\mathbf{H}/\overline{\mathbf{H}}/\underline{\mathbf{H}})$$

Multi-Head Attention

Key Value Query

$$K^h = \mathbf{H}W_k^h, \;\; V^h = \mathbf{H}W_v^h, \;\; Q^h = \mathbf{H}W_q^h$$

H : number of heads

$W_k^h, W_v^h, W_q^h \in \mathbb{R}^{d\times d_k}$ $h = 1, \dots, H$

d_k : dimension of key, value, query in each head

Multi-Head Attention Output

$$AO^h = \text{Atten}(Q^h, K^h, V^h) = \text{softmax}\left(\frac{Q^h (K^h)^T}{\sqrt{d_k}}\right) V^h$$

$$MAO = Concat\left(\{AO^h\}_{h=1}^{H}\right) W^A$$

W^A : Trainable projection matrix

Part (b) Mix-up Frequency Representations

$$\tilde{\mathbf{R}} = Concat(\mathbf{R}, \overline{\mathbf{R}}, \underline{\mathbf{R}})$$

Prediction Head

$$\hat{y} = MLP(\tilde{\mathbf{R}})$$

$$\mathcal{L}_{rul} = \mathbb{E}\left[\sum_{i=1}^{P} \|y_i - \hat{y}_i\|_2\right]$$

Masked Autoencoder Reconstruction Auxiliary Learning (MARAL)

Part (c) Masking and Tokenize

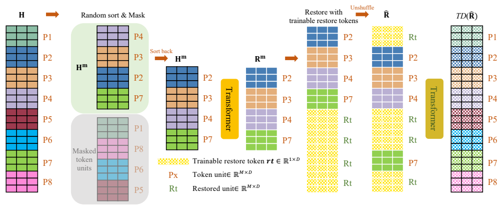


Figure 2. Masking and Tokenize Strategy

Masking and Tokenize

$$\mathbf{H}^m = Mask(\mathbf{H}) \in \mathbb{R}^{(P-kM)\times D}$$

Feature Projection

$$\mathbf{R}^m = TE(\mathbf{H}^m) \in \mathbb{R}^{(P-kM)\times D}$$

Restored Unit Generator

$$\mathbf{Rt} = Concat(\{\mathbf{rt_m}\}_{m=1}^{M}), \;\; \mathbf{rt_1} = \mathbf{rt_2} = \dots = \mathbf{rt_M}$$

Restoration

$$\hat{\mathbf{R}} = Restore(Concat(\mathbf{R}^m, \{\mathbf{Rt_i}\}_{i=1}^{k})) \in \mathbb{R}^{P\times D}$$

Part (d) Signal Reconstruction

$$\hat{\mathbf{X}} = FC(TD(\hat{\mathbf{R}})) \in \mathbb{R}^{P\times J}$$

Reconstruction Loss Function: Cosine Similarity-based Loss Function

$$\mathcal{L}_{aux} = \frac{1}{B}\sum_{b=0}^{B}\sum_{p=0}^{P}\left(1 - \frac{\mathbf{x}_p^b \cdot \hat{\mathbf{x}}_p^b}{\max\left(\|\mathbf{x}_p^b\|_2 \cdot \|\hat{\mathbf{x}}_p^b\|_2, \epsilon\right)}\right)$$

RUL Labeled Data

$$\mathcal{L} = \lambda_1 \mathcal{L}_{rul} + \lambda_2 \mathcal{L}_{aux}$$

RUL Labeled Data + Unlabeled Data (domain unrestricted, given fake label)

$$\mathcal{L} = \lambda_1 \mathcal{L}_{rul}^{l} + \lambda_2 \mathcal{L}_{aux}^{l} + \lambda_3 \mathcal{L}_{rul}^{ul} + \lambda_4 \mathcal{L}_{aux}^{ul}$$

Experimental Setup and Results

1. Data Preparation

Dataset	Operating conditions	Fault modes	Train size (nr. of engines)	Test size (nr. of engines)
FD001	1	1	100	100
FD002	6	1	260	259
FD003	1	2	100	100
FD004	6	2	248	249

Table 1. Basic Information of C-MAPSS

3. Fully-supervised model Ablation Study

Model	LSTM	FEMM	MARAL	RMSE	Score
I				12.05	218.07
II	✓			12.61	257.26
III		✓		11.33	195.88
IV	✓	✓		12.63	280.84
V			✓	11.62	212.62
VI	✓		✓	11.13	184.57
VIII		✓	✓	12.08	259.14
IX	✓	✓	✓	**10.79**	**179.10**

* Both RMSE and Score are the lower the better
* The **bold** results mark the best results.

Table 3. Results of ablation study on FD001

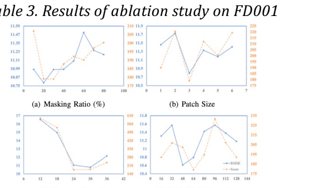


Figure 4. The effects of the (a)masking ratio, (b)patch size, (c)sequence length, and (d)embedded dimension on the prognostic performance for the training process on FD001. The blue line is RMSE and the orange dot line is the score.

2. Fully-supervised Performance Comparisons (only RUL dataset)

Method	FD001		FD002		FD003		FD004	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
SVR [22]	20.96	1380	42.00	590000	21.05	1600	45.35	371000
CNN [22]	18.45	1299	30.29	13600	19.82	1600	29.16	7890
Deep LSTM [23]	16.14	388	24.49	4450	16.18	852	28.17	5550
Auto-Encoder [24]	13.58	228	19.59	2650	19.16	1727	22.15	2901
GCU-Transformer [14]	11.27	-	22.81	-	11.42	-	24.86	-
RVE [25]	13.42	323.82	14.92	1379.17	12.51	256.36	16.37	1845.99
MCTAN [26]	11.69	189.04	16.41	**446.89**	**10.72**	252.30	17.36	1722.22
Ours	**10.79**	**179.10**	**13.03**	728.73	11.59	**206.56**	**14.61**	**1038.54**

Table 2. Evaluation metrics of different approaches for RUL estimation on C-MAPSS datasets

4. Results: SSL on Low-data Settings Performance Comparisons

(Unlabelled RUL dataset: FD002 (only take the first 50%), Cross-domain dataset: SleepEEG)

Model \ Percentage (%)	10		20		30		40		50	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
RVE	61.05	82554.37	48.32	17880.00	22.74	2483.60	22.70	2492.11	20.15	1293.37
GCU-Transformer	75.22	148110.54	32.91	2290.16	16.37	377.73	15.85	294.48	13.36	282.28
Ours: FS	57.06	25853.26	34.52	2634.76	20.54	579.91	15.10	287.03	12.88	217.44
Ours: FS+SSL with RUL Data	**16.96**	**492.68**	15.25	357.43	**13.52**	290.78	14.03	263.60	**12.49**	207.85
Δ	−40.10	−25360.58	−19.27	−2277.33	−7.12	−289.13	−1.07	−23.43	−0.39	−9.59
Ours: FS+SSL with Cross-Domain Data	27.67	1601.68	**14.52**	**273.00**	13.64	**223.32**	**13.92**	**265.90**	12.71	**207.40**
Δ	−29.39	−24251.58	−20.00	−2361.76	−7.00	−356.59	−1.18	−21.13	−0.17	−10.04

Model \ Percentage (%)	60		70		80		90		100	
	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score	RMSE	Score
RVE	20.12	1289.82	16.22	560.34	16.24	581.25	15.37	492.53	15.19	452.22
GCU-Transformer	12.68	225.36	12.16	191.15	11.73	187.70	11.93	222.01	11.40	179.96
Ours: FS	12.13	199.64	11.25	159.12	10.73	169.07	10.35	150.42	10.79	179.10
Ours: FS+SSL with RUL Data	**10.83**	**166.41**	**10.24**	150.90	**9.77**	**129.06**	9.99	**134.04**	10.43	**146.20**
Δ	−1.3	−33.23	−1.01	−8.22	−0.96	−40.01	−0.36	−16.38	−0.36	−32.90
Ours: FS+SSL with Cross-Domain Data	11.63	180.48	10.50	**149.52**	10.32	160.33	**9.76**	144.76	**10.36**	161.79
Δ	−0.50	−19.16	−0.75	−9.6	−0.41	−8.74	−0.59	−5.66	−0.34	−17.31

Table 4. Evaluation metrics of semi-supervised learning (SSL) with different percentage of FD001

5. Results: SSL on Low-data Settings Weights Sensitivity Experiments

λ_3 \ λ_4	0	0.1	0.15	0.2	0.25	0.3
10%	$mean = 18.92$ $std = 1.40$					
0	57.06	60.86	61.00	62.27	58.01	64.50
0.05	19.09	17.60	17.62	**16.96**	18.47	17.90
0.1	21.43	18.37	18.75	18.08	17.91	18.28
0.15	23.18	18.51	18.74	18.67	19.03	18.80
0.2	21.96	18.53	19.37	18.79	18.80	19.30
20%	$mean = 15.87$ $std = 0.31$					
0	34.52	36.41	35.53	35.74	37.53	39.63
0.05	16.23	15.64	15.65	**15.25**	15.37	15.68
0.1	15.70	15.97	15.93	15.93	16.13	16.11
0.15	16.69	16.19	15.76	15.81	15.62	15.90
0.2	16.06	15.63	16.17	16.14	15.63	15.92

* λ_1 and λ_2 are fixed to be 1 and 1.05
* std and $mean$ are calculated without zero λ_3 cases.

Table 5. SSL loss weights sensitivity experiments results (RMSE) with RUL unlabelled

λ_3 \ λ_4	0	0.002	0.005	0.007	0.01	0.02
10%	$mean = 37.98$ $std = 6.45$					
0	57.06	50.37	58.19	61.77	56.57	60.85
0.002	30.58	37.87	33.45	37.47	29.15	40.77
0.005	37.55	28.80	37.57	45.15	45.64	49.90
0.007	35.71	**27.67**	31.25	39.55	37.63	42.69
0.01	32.69	34.19	37.05	43.34	49.76	46.19
20%	$mean = 19.27$ $std = 4.04$					
0	34.52	31.80	37.55	33.36	33.44	38.95
0.002	23.06	16.64	**14.52**	16.73	24.14	20.64
0.005	28.99	22.39	14.93	14.60	19.08	16.42
0.007	23.26	18.42	14.92	19.41	15.79	22.19
0.01	26.26	19.39	14.86	17.28	16.33	22.39

λ_1 and λ_2 are fixed to be 1 and 1.05
std and $mean$ are calculated without zero λ_3 cases.

Table 6. SSL loss weights sensitivity experiments results (RMSE) with cross-domain unlabelled

Conclusion

- The **LSTM temporal projection** layer effectively captures **long sequence time-series information.** The **FEMM module** enhances feature extraction using **frequency-domain decomposition.**
- The **MARAL** addresses data scarcity and boosts model performance, applicable to **unrestricted domain datasets.**
- On the C-MAPSS dataset, particularly in FD004, demonstrated our method's superiority with a **10.75% RMSE drop,** and a **39.69% score reduction** compared to the second-best results. By introducing unrestricted domain data for auxiliary learning, using just **10% labeled data**, there was an **impressive 98.09% score drop** and a **70.28% RMSE decline** compared to the fully supervised model.