Electronic supplementary materials

For https://doi.org/10.1631/jzus.A2500300

PL-HLANet: a semi-supervised approach for tunnel boring machine disc cutter wear prediction

Zhaoyang LI, Wei TANG, Xinyuan WANG, Huxiu XU, Huayong YANG, Jun ZOU

State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou 310058, China

Section S1 Detailed Descriptions of Input and Output Variables

This section provides supplementary details regarding the dataset and specific feature calculation methods used in this study, corresponding to the overview presented in Section 2.3 of the main paper.

Table S1 presents a comprehensive list of the 11 input features and 4 output targets. Eqs. S1-S3 define the calculation methods for the composite geological indices (FPI, TPI).

$$\mathrm{FPI} = \frac{F}{P} \tag{S1}$$

$$\mathrm{TPI} = \frac{T}{P} \tag{S2}$$

$$P = \frac{v}{n} \tag{S3}$$

Table S1 Detailed list of the 11 input features and 4 output targets used in this study.

No.	Parameter	Unit	Mean Value	Symbol
1	Penetration rate	mm/min	40.68	X_1
2	CHD rotational speed	rpm	4.48	X_2
3	Thrust	kN	20110.95	X_3
4	CHD torque	kN.m	3188.15	X_4
5	Penetration	mm/r	9.66	X_5
6	FPI	KN/(mm/r)	38.57	X_6
7	TPI	KN·m/(mm/r)	5.56	X_7
8	CHD water flow	L/min	176.54	X_8
9	Gripper pressure	bar	262.31	X_9
10	TBM roll angle	mm	-35.23	X_{11}
11	Left shoe longitudinal deviation angle	°	-1.59	X_{12}
12	C29 cutter wear rate	mm/h	0.19	y_1
13	C54 cutter wear rate	mm/h	0.38	y_2
14	C60 cutter wear rate	mm/h	0.63	y_3
15	C65 cutter wear rate	mm/h	0.51	y_4

Section S2 Hyperparameter settings for baseline models

This section provides the detailed hyperparameter settings for all baseline models used in the comparative performance evaluation presented in Section 4.2 of the main paper. To ensure a fair and robust comparison, all baseline models were carefully tuned to achieve their reasonable optimal performance. The specific configurations, which were used for all our comparative experiments, are detailed in Table S1.

Table S2 Hyperparameter settings for comparative models.

Model	Key Parameters	Optimizer	Learning Rate	Batch Size	Epochs
LSTM	Layers=6, Hidden Dims=256, Dropout=0.2	Adam	2e-4	64	200
GRU	Layers=6, Hidden Dims=256, Dropout=0.2	Adam	2e-4	64	200
Transformer encoder	Layers=6, Heads=4, d_model=128, Dropout=0.2	Adam	2e-4	64	200
SVR	Kernel=RBF, C=1.0, γ='scale'	-	-	-	-
Mean Teacher	Backbone=Residual CNN (3 Blocks, Dropout=0.2), α=0.9999	Adam	2e-4	64	200
FixMatch	Backbone=Residual CNN (3 Blocks, Dropout=0.2), λ_u=0.25, τ=0.15	Adam	2e-4	64	200
PL-HLANet	TCN (Blocks=3, Channels=64, Kernel=7, Dropout=0.2), Bi-LSTM (Layers=1, Hidden=128)	Adam	2e-4	64	200

**Note: All deep learning models were trained using the Adam optimizer with a learning rate of 2e-4 and a batch size of 64. An early stopping mechanism with a patience of 20 epochs was applied.*

Section S3 Ablation study on the PL module

This section provides the detailed visual evidence for the core ablation study on the effectiveness of the Pseudo-Labeling (PL) module. While the quantitative contribution of the PL module is summarized in Table 7 of the main manuscript, the following Fig. S1 presents the intuitive prediction curve comparisons for all cutters to more comprehensively illustrate the module's role in improving the model's tracking capability.

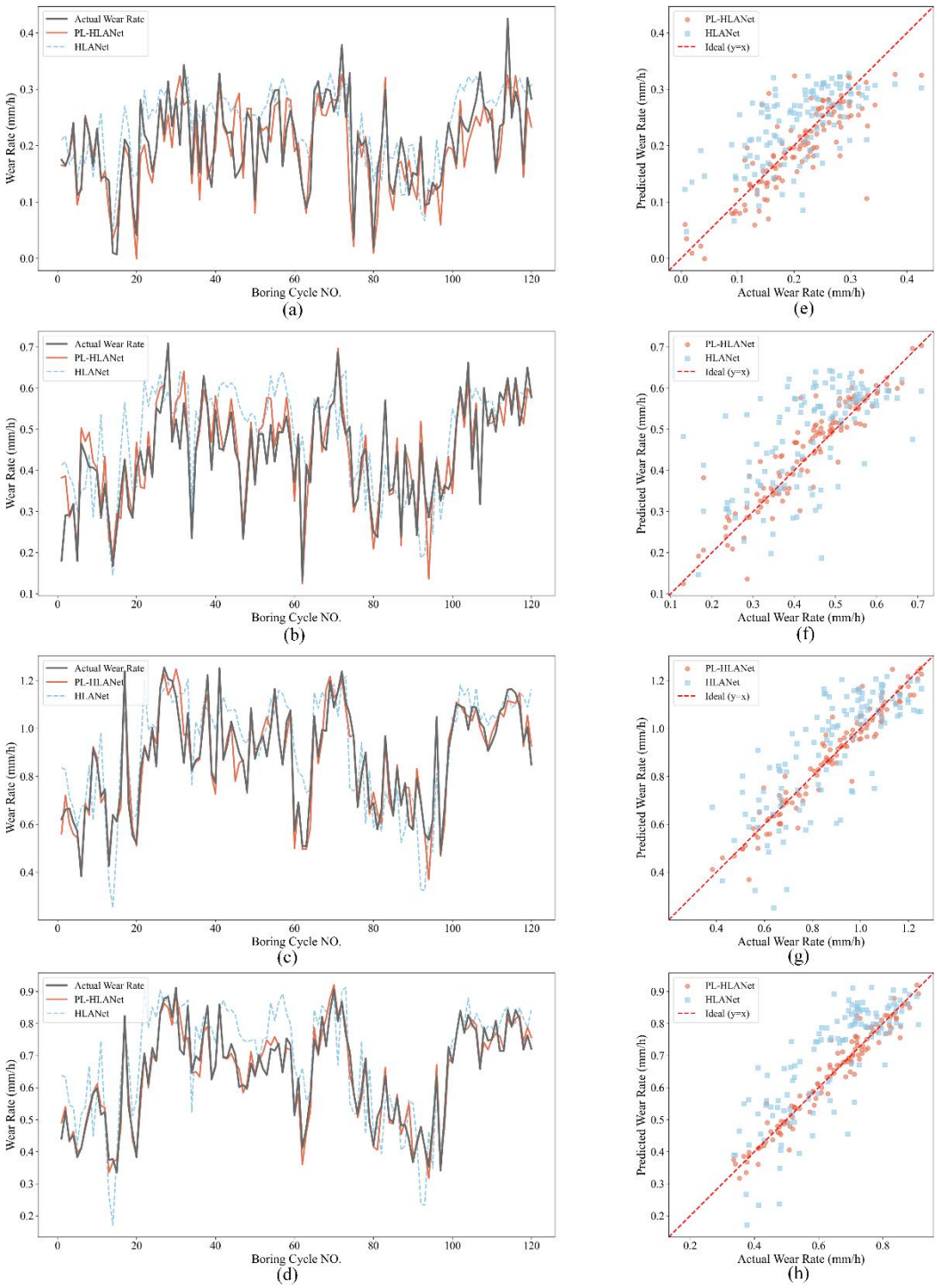

Fig. S1 Ablation study on the PL module. The left column (a-d) displays the line plots of instantaneous wear rates, while the right column (e-h) displays the corresponding scatter plots of cumulative wear. From top to bottom, rows correspond to cutters C29, C54, C60, and C65, respectively.

Section S4 Ablation study on the HLANet module

This section aims to supplement the ablation study results of the HLANet module's internal components with detailed visualizations. While the quantitative contributions of each component are demonstrated in Table 8 of the main manuscript, the following Fig. S2 intuitively reveals the impact on the model's predictive behavior when different components (TCN, Bi-LSTM, Attention) are removed by showing the prediction curve comparisons on a representative cutter.

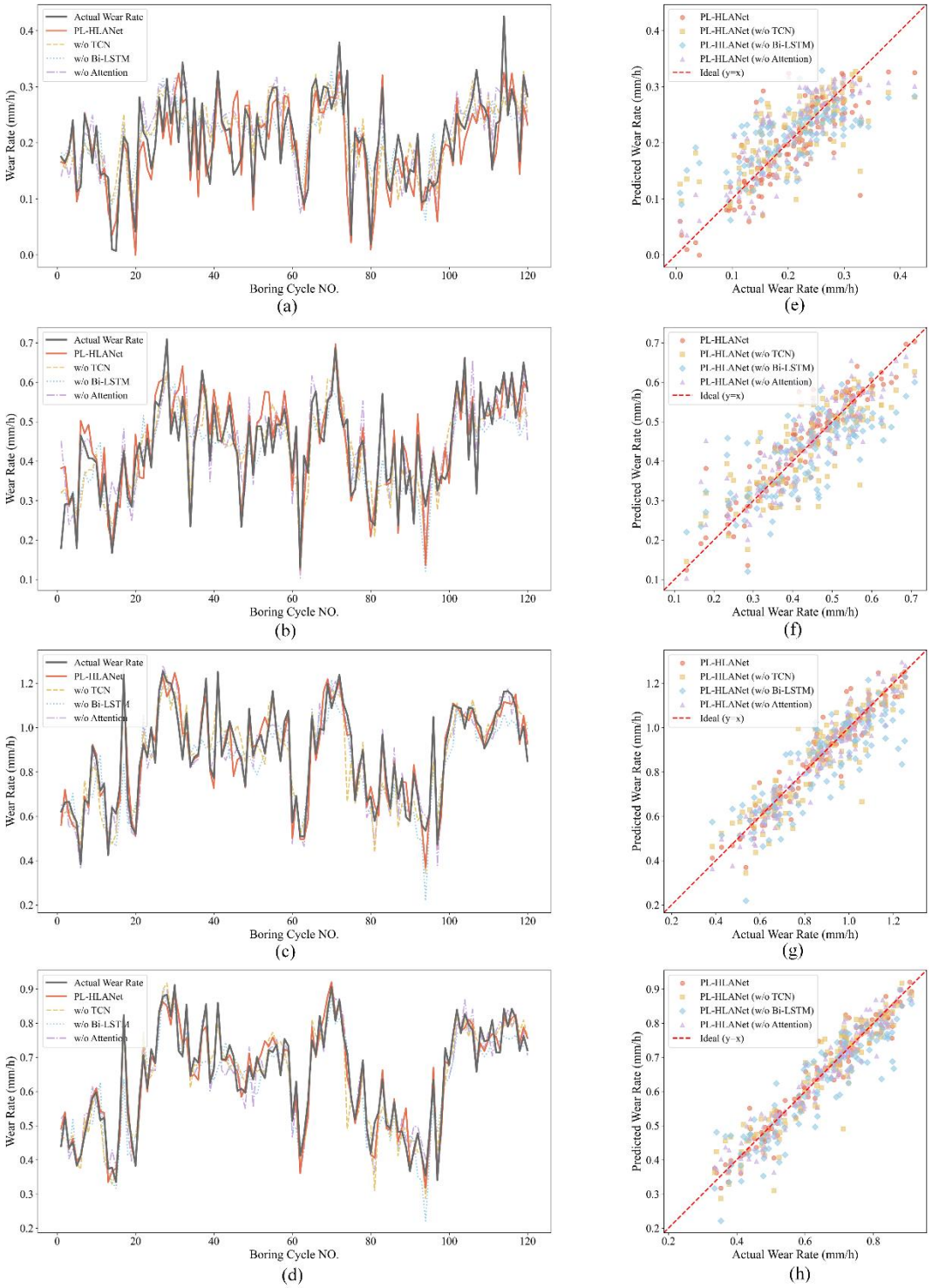

Fig. S2 Ablation study on the HLANet module. The left column (a-d) displays the line plots of instantaneous wear rates, while the right column (e-h) displays the corresponding scatter plots of

cumulative wear. From top to bottom, rows correspond to cutters C29, C54, C60, and C65, respectively.

Section S5 Ablation study on the Bi-LSTM architecture

To systematically evaluate the impact of Bi-LSTM network depth on model performance, we conducted an ablation study. We rigorously compared the predictive efficacy of models incorporating single-layer, dual-layer, and triple-layer Bi-LSTM configurations on the test set. Except for the network depth, all other parameters and training configurations were kept identical across all models.

The results, presented in Table S3, clearly indicate that increasing the number of Bi-LSTM layers did not yield a significant improvement in performance. Instead, it led to a multiplicative increase in the number of parameters, substantially raising the model's complexity and computational overhead. Specifically, the dual-layer model performed comparably to the single-layer version, while the triple-layer configuration led to a distinct decline in predictive accuracy. While this outcome may seem counter-intuitive to the common wisdom that deeper networks can extract temporal features more thoroughly, it is well-explained by the synergistic role that Bi-LSTM plays within our unique HLANet architecture.

In HLANet, feature extraction is not the sole responsibility of the Bi-LSTM but is rather a relay process involving a collaboration between the TCN, Bi-LSTM, and Attention modules. The preceding TCN module has already efficiently extracted critical local temporal features, while the subsequent Attention layer is tasked with intelligent focusing. This allows the intermediate Bi-LSTM to concentrate on its core task: enriching the high-level feature sequence refined by the TCN with rich, long-term, bidirectional contextual information. For this well-defined task, a single Bi-LSTM layer is entirely sufficient.

Therefore, comprehensively considering both predictive accuracy and model lightweighting (parameter efficiency), we conclude that a single-layer Bi-LSTM is the choice that achieves the best cost-effectiveness within the HLANet architecture, representing a deliberate and well-weighed design choice.

Table S3 Ablation study results of the Bi-LSTM architecture.

Model	Bi-LSTM Layers	R^2	MAE	RMSE	Parameters
PL-HLANet-L1	1	0.8944	0.0307	0.0437	376719
PL-HLANet-L2	2	0.8961	0.0326	0.0454	771983
PL-HLANet-L3	3	0.8613	0.0411	0.0540	1167247

Section S6 Analysis of the consistency threshold

The consistency threshold ε is a core hyperparameter in our multi-view pseudo-label generation module. To systematically investigate its impact and identify the optimal value, we conducted a two-stage parameter sensitivity analysis.

(1) Effect of ε on the pseudo-label generation process

First, we explored the trade-off between the "quantity" and "quality" of the pseudo-labels generated under different ε values. We measured quantity by the number of common pseudo-labels generated and used the "Average Standard Deviation (Avg. Std. Dev.)" as an inverse indicator of quality.

The results, summarized in Table S4, show that as ε increases, the number of generated pseudo-labels grows substantially, but at the cost of declining quality, as indicated by the increasing Avg. Std. Dev.. This stage provided several candidate pseudo-label sets with different quantity-quality profiles for final performance validation.

Table S4 Impact of different ε values on the quantity and quality of pseudo-labels.

ε	# Pseudo- Labels	Avg. Std. Dev.
0.02	39	0.0084
0.05	698	0.0190
0.10	1998	0.0289
0.15	2361	0.0327
0.30	2946	0.0385
0.50	2982	0.0390
1.00	2982	0.0391

(2) Effect of ε on final model performance

To determine which pseudo-label combination yields the best predictive model, we used the four sets of pseudo-labels (generated with ε = 0.05, 0.10, 0.15, and 0.30), combined them with the original labeled data, and trained our HLANet model from scratch for each case. All models were evaluated on the same independent test set, with the results presented in Table S5 and Fig. S3.

Table S5 Performance comparison of HLANet models trained with different ε values.

ε	R^2	MAE	RMSE
0.05	0.5595	0.0635	0.0866
0.10	0.8216	0.0364	0.0551
0.15	0.8944	0.0307	0.0437
0.30	0.8531	0.0326	0.0513

A comprehensive analysis of both stages leads to the conclusion that ε = 0.15 is the optimal threshold in this study.

Interestingly, the model trained with the highest-quality pseudo-labels (lowest Avg. Std. Dev. at ε=0.05) did not yield the best performance. As ε was relaxed from 0.05 to 0.15, the number of pseudo-labels increased substantially from 698 to 2361, despite a slight decrease in their average quality. This result suggests that for a complex deep model like HLANet, once the pseudo-label quality reaches a good enough level, having a larger and more diverse set of pseudo-labels plays a more critical role in enabling sufficient training, regularization, and enhancing generalization ability. When ε was further relaxed to 0.30, the model's performance began to decline due to the introduction of too many low-quality labels.

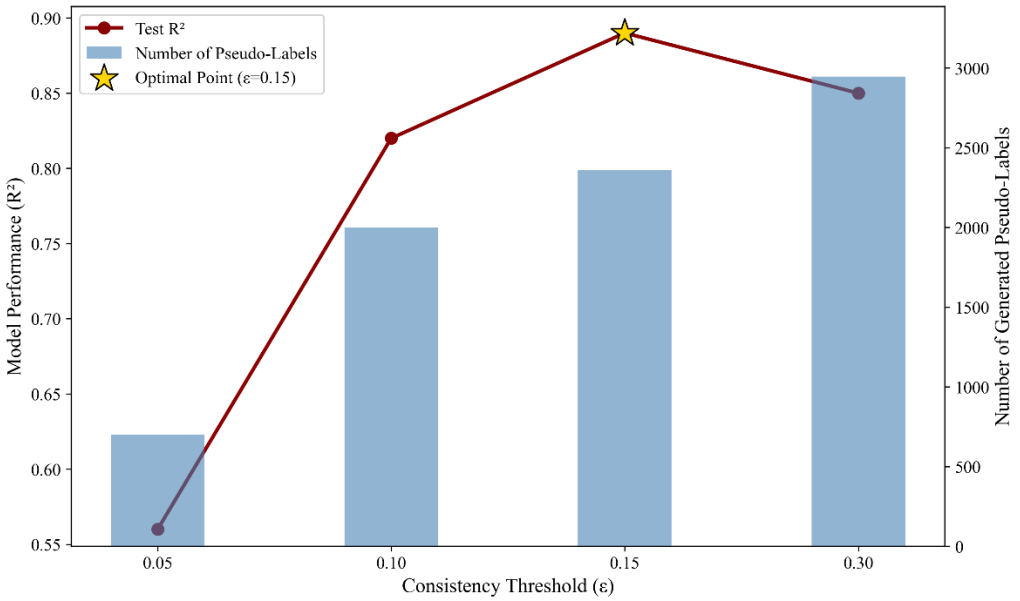

Fig. S3 Impact of ε on pseudo-label quantity and model prediction performance.