Similarity-based Framework for Remaining Useful Life Prediction in Semiconductor Manufacturing Equipment

Takanobu Minami[1,2] and Lee Jay[1]

[1]*Center for Industrial Artificial Intelligence, Department of Mechanical Engineering, University of Maryland, College Park, MD, USA*
minamitu@umd.edu, leejay@umd.edu

[2]*Komatsu Ltd., Tokyo, Japan*

ABSTRACT

Semiconductor manufacturing equipment poses a demanding setting for Remaining Useful Life (RUL) prediction due to regime-rich operation, heterogeneous sensors, and practical requirements for interpretability and rapid deployment. This work examines a similarity-based alternative on the PHM2018 ion mill etching (IME) dataset, using Tool 01_M02 and the F1 fault mode as a case study. The pipeline comprises four stages: (i) regime-aware normalization, where operating-condition features are standardized and clustered on the training partition and sensor channels are expressed as cluster-wise z-scores; (ii) supervised health-indicator construction via weighted Ridge regression using correlation-selected channels with time-normalized derivatives; (iii) monotonic calibration that blends the indicator with a simple time prior and fits an isotonic mapping; and (iv) neighbor-based estimation in which a quadratic is fitted to each training trajectory and test-time RUL is inferred by distance-weighted aggregation across nearest trajectories.

Evaluation follows a leave-one-cycle-out (LOCO) protocol under a common IME preprocessing setup (active-state filtering and tail restriction at RUL $\leq$ 5000). On F1, the approach attains an average RMSE of 453.53, outperforming a state-of-the-art baseline (ATCN-LSTM, 597.35). One-factor sensitivity analyses show consistent trends: performance improves as the number of selected features decreases; shallow optima appear for the derivative window and smoothing window; smaller blend factors are generally favorable; and the late-life weighting exponent has minor influence. The computational profile is lightweight: the full 62-fold LOCO evaluation completes in about one minute on a CPU-only workstation (≈1.0 s per fold), facilitating rapid iteration and deployment. These results indicate that a similarity-based framework which centered on regime-aware indicator design, monotonic calibration, and neighbor matching can deliver competitive accuracy while preserving transparency and practical efficiency for industrial PHM.

1. INTRODUCTION

Semiconductor manufacturing equipment plays a pivotal role in sustaining progress in integrated circuit and photovoltaic industries, yet failure prognostics remains challenging due to regime-rich operation, heterogeneous sensors, and complex subsystem interactions. Downtime from unexpected failures or scheduled maintenance can incur substantial losses in high-volume production.

Prognostics and Health Management (PHM) addresses these challenges by estimating Remaining Useful Life (RUL) to support condition-based maintenance. Over the past decade, data-driven approaches, especially deep models such as LSTM and TCN, have achieved strong results on benchmark datasets. However, they often require large, labeled datasets and significant computational resources, and their limited interpretability can hinder industrial adoption.

As an alternative, similarity-based prediction estimates RUL by comparing a target trajectory with previously observed ones. This family of methods is appealing for semiconductor equipment, where fault data are relatively scarce, operating conditions vary across runs, and traceability of predictions is important for industrial adoption. By focusing on relative similarity rather than learning a global mapping, similarity-based approaches can remain data-efficient and transparent while aligning with standard evaluation practice.

In this study, a similarity-based methodology is evaluated on the PHM2018 ion mill etching (IME) dataset as a case study. The pipeline integrates regime-aware normalization, supervised health-indicator construction, monotonic calibration, and neighbor-based estimation, and is assessed under leave-one-cycle-out (LOCO) evaluation to enable

direct comparison with deep baselines. Results show that the approach outperforms a state-of-the-art deep model, while preserving computational efficiency on CPU-only hardware, positioning it as a lightweight and interpretable alternative for industrial RUL prediction in complex manufacturing systems.

2. RELATED WORK

Recent surveys on PHM identify RUL prediction as a core enabler of condition-based maintenance and document an ongoing shift from physics-based modeling to data-driven and deep learning approaches, while highlighting persistent challenges in data scarcity, operating-condition shift, interpretability, benchmarking practice, and deployment cost (Ferreira & Gonçalves, 2022; Berghout & Benbouzid, 2022; Wu et al., 2024; Liu et al., 2025). These reviews commonly motivate hybrid or lightweight alternatives that balance accuracy with transparency and practical efficiency.

Within manufacturing, semiconductor equipment has emerged as a representative and demanding RUL setting due to regime-rich operation and heterogeneous sensor suites. A concrete line of work has coalesced around the PHM2018 IME dataset: recurrent baselines for the flowcool subsystem have been explored (Wu et al., 2021); hybrid TCN–LSTM with attention has been proposed to capture both local and long-range dependencies (Hsu et al., 2022); transfer-learning pipelines align tools or conditions prior to fine-tuning (Liu et al., 2021); and Transformer-based variants introduce channel re-weighting or multi-scale, multi-branch temporal encoders (Yuan & Wang, 2023; Yuan & Wang, 2024). Within this stream, ATCN-LSTM (Darwish, 2024) is frequently adopted as a strong deep baseline and serves as the principal point of reference for comparison in the present study.

In parallel, similarity-based prediction (SBP) has been surveyed as an interpretable and lightweight alternative in PHM: a degradation indicator is constructed, similarity to historical trajectories is evaluated, and RUL is inferred via neighbor aggregation (Xue et al., 2022). Foundationally, trajectory SBP by Wang et al. (2008) and its extended dissertation (Wang, 2010) formalized trajectory-level matching and influenced numerous variants. This lineage is followed here for IME by combining regime-aware indicator design and neighbor-based estimation.

Taken together, prior work shows that IME RUL has been dominated by deep neural pipelines whereas similarity-based approaches have been less examined for semiconductor equipment despite their interpretability and low compute. This gap frames the present study, which assesses a similarity-based methodology on PHM2018 IME dataset and contrasts it with a strong deep baseline (ATCN-LSTM) under a common evaluation setup (Darwish, 2024).

3. METHODOLOGY

The proposed approach consists of four main stages: (i) regime-based normalization, (ii) supervised health indicator (HI) construction, (iii) monotonic calibration, and (iv) similarity-based RUL estimation using polynomial trajectory fitting and neighbor search. A schematic overview of the framework is shown in Figure 1.

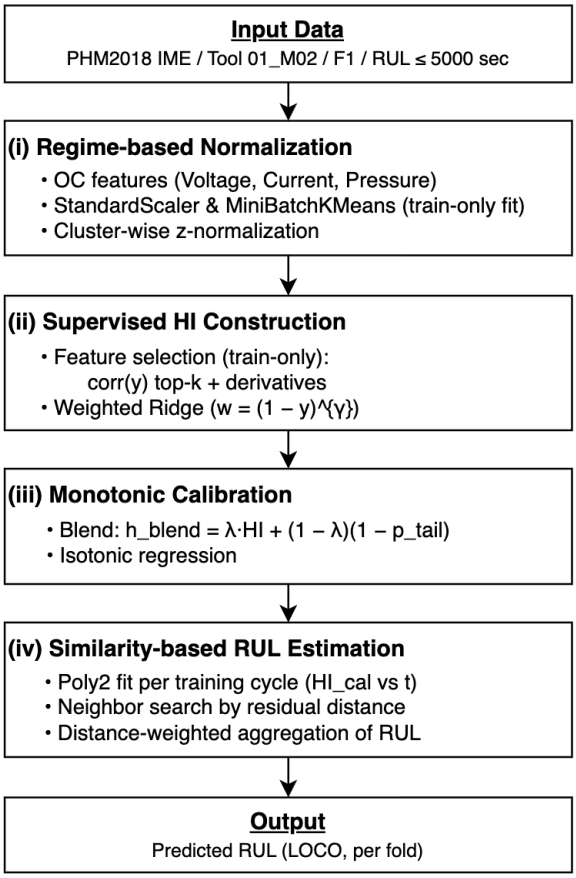


Figure 1. Overview of the proposed similarity-based RUL prediction methodology

3.1. Regime-based Normalization

To mitigate variability arising from heterogeneous operating conditions, operating-condition (OC) features (e.g., beam voltage, beam current, flowcool pressure) are standardized and clustered on the training partition of each fold.

Let $x_i \in \mathbb{R}^m$ denote the OC feature vector for sample i. Standardization and clustering are performed as

$$z_i = \frac{x_i - \mu}{\sigma}, \quad c_i = \arg\min_{k \in \{1,\dots,K\}} || z_i - \mu_k ||_2 \qquad (1)$$

where μ, σ are the training mean and standard deviation of the OC features and μ_k is the centroid of cluster k obtained by MiniBatchKMeans. The learned scaler and centroids are then applied to both training and test samples within the fold.

Each sample is assigned a regime label c_i.

Given the regime labelc_i, each sensor channel s is normalized in a cluster-wise manner to capture deviations relative to its regime:

$$s_{i,zn} = \frac{s_i - \mu_{s,c_i}}{\sigma_{s,c_i} + \epsilon} \quad (2)$$

with μ_{S,c_i} and σ_{S,c_i} computed on the training partition for regime c_i, and small constant $\epsilon > 0$ added to avoid division by zero when within-regime variance is negligible. This transformation reduces between-regime variance while preserving regime-consistent degradation trends, providing a common scale on which feature selection and HI construction (Section 3.2) operate effectively.

3.2. Feature Selection and Health Indicator Construction

A supervised target is defined on the tail region as

$$y_i = \frac{RUL_i}{RUL_{max}} \in [0, 1] \quad (3)$$

where RUL_{max} is specified in Section 4.2. From cluster-wise normalized sensor channels $s_{i,zn}$ (Section 3.1), the top-k features are selected on the training partition using Pearson correlation with y:

$$\rho(s, y) = \frac{\mathrm{Cov}(s, y)}{\sigma_s \sigma_y}, \mathcal{F} = \{s\text{:top-}k \text{ by } |\rho(s, y)|\} \quad (4)$$

To capture short-horizon change, finite-difference derivatives are added for each $s \in \mathcal{F}$:

$$s_i^{(d)} = \frac{s_i - s_{i-d}}{t_i - t_{i-d}} \quad (5)$$

where t_i denotes elapsed time in seconds within a cycle and d denotes a window of length; when $t_i - t_{i-d} \leq 0$ the derivative is set to 0 and optionally clipped to a bounded range for numerical robustness. Time-progress covariates p_{tail} and p_{tail}^2 are further included. The resulting feature vector is

$$z_i = \left[s_{i,\mathrm{zn}}\,(s! \in !\mathcal{F})\,;\; s_i^{(d)}\,(s! \in !\mathcal{F})\,;\; p_{\mathrm{tail},i},\; p_{\mathrm{tail},i}^2 \right] \in R^m \quad (6)$$

A weighted Ridge model is then fitted to construct a health indicator (HI). Let $\tilde{z}_i$ denote standardized features (mean–variance scaling on the training partition). The HI prediction is

$$\hat{h}_i = f(z_i) = w^\top \tilde{z}_i \quad (7)$$

with parameters obtained by

$$w = \arg\min_w \sum_{i \in \mathcal{D}_{train}} w_i (y_i - w^\top \tilde{z}_i)^2 + \alpha ||w||_2^2 \quad (8)$$

Here, $w_i = (1 - y_i)^\gamma$ assigns greater importance to late-life samples, α is the Ridge regularization parameter, and γ controls the degree of weighting. The resulting HI trajectory reflects degradation progression in a data-driven yet interpretable manner.

3.3. Monotonic Calibration

To promote a monotonic degradation profile, the predicted HI is blended with a simple time-based prior and then calibrated by isotonic regression. The blended indicator is defined as

$$h_i^{\mathrm{blend}} = \lambda \hat{h}_i + (1 - \lambda)\left(1 - p_{\mathrm{tail},i}\right) \quad (9)$$

where $\lambda \in [0, 1]$ controls the contribution of the model prediction versus the prior. An isotonic mapping $g(\cdot)$ is then fitted on the training partition of each fold and applied to both training and test samples within the fold:

$$g = \arg\min_{g \in \mathrm{Isotonic}} \sum_i \left(y_i - g\left(h_i^{\mathrm{blend}}\right) \right)^2 \quad (10)$$

$$h_i^{\mathrm{cal}} = g\left(h_i^{\mathrm{blend}}\right) \quad (11)$$

The learned mapping enforces a monotone relationship consistent with the decreasing target in the tail region, yielding a calibrated trajectory h^{cal} that is smoother and aligned with expected degradation.

3.4. Similarity-based RUL Estimation

For each fold, the calibrated HI trajectory of every training cycle is approximated on the normalized tail time $t \in [0,1]$ by a quadratic

$$h_c^{poly}(t) = a_c t^2 + b_c t + c_c \quad (12)$$

fitted on that training cycle's $h^{cal}(t)$. Given a test trajectory observed up to t^*, let $T^* \subset [0, t^*]$ be the sampled grid. The dissimilarity between the test and a training cycle c is measured by the mean absolute residual

$$d(c) = \frac{1}{|T^*|} \sum_{t \in T^*} \left| h_{\mathrm{test}}^{\mathrm{cal}}(t) - h_c^{\mathrm{poly}}(t) \right| \quad (13)$$

The k training cycles with the smallest $d(c)$ form the neighbor set $\mathcal{N}$ (if fewer than k candidates are available, all available neighbors are used).

To map the test end-point to the neighbor's time scale, $t'_c \subset [0, 1]$ is obtained by solving

$$a_c t'^2 + b_c t' + c_c = h_{\mathrm{test}}^{\mathrm{cal}}(t^*) \quad (14)$$

If two admissible real roots exist in $[0, 1]$, the one closest to t^* is selected. If no admissible root exists, that neighbor is skipped. The neighbor-specific RUL is then

$$\widehat{RUL_c} = (1 - t'_c) \cdot RUL_{max} \quad (15)$$

Finally, the prediction is aggregated with distance-based weights

$$\widehat{RUL} = \frac{\sum_{c\in\mathcal{N}}(d(c)+\epsilon)^{-p}\ \widehat{RUL_c}}{\sum_{c\in\mathcal{N}}(d(c)+\epsilon)^{-p}} \tag{16}$$

where N is the set of neighbors, $\epsilon > 0$ safeguards numerical stability and $p > 0$ controls the decay with distance. All quantities used for fitting (quadratic coefficients and distances) are computed on training cycles within the fold, and the resulting prediction is evaluated on the held-out test cycle.

4. EXPERIMENTAL SETUP

This section outlines the dataset, preprocessing pipeline, evaluation protocol, hyperparameter configuration, and computational environment used in the study.

4.1. Dataset

Experiments were conducted on the PHM2018 IME dataset, which contains multivariate condition-monitoring records from twenty tools. The dataset provides 24 feature variables sampled at 4-second intervals, comprising five categorical attributes (e.g., wafer ID, tool ID, recipe) and nineteen numerical sensor channels (e.g., voltage, current, pressure, flow). Three failure types are annotated: F1 (Flowcool Pressure Dropped Below Limit), F2 (Flowcool Pressure Too High Check Flowcool Pump), and F3 (Flowcool Leak).

In this work, the scope is restricted to Tool 01_M02 and the F1 fault mode as a case study, consistent with prior work. Run-to-failure traces were consolidated into cycles and RUL labels were assigned accordingly.

4.2. Preprocessing

Raw records were prepared through a sequence of standard steps prior to model training. First, only samples corresponding to the active operating state were retained (FixtureShutterPosition = 1). Each run-to-failure trace was then consolidated into cycles, and an elapsed time axis was computed per cycle relative to its start to provide a consistent temporal reference.

Next, the dataset was restricted to the tail region, retaining only samples with RUL ≤ 5000; observations with RUL > 5000 were regarded as normal operating conditions and excluded, following prior practice on this dataset. This restriction emphasizes degradation progression and provides a common prediction target across experiments.

Finally, regime-based normalization was applied to mitigate variability across operating conditions. Operational-condition features (e.g. beam voltage, beam current, and flowcool pressure) were standardized and clustered on the training partition of each fold (MiniBatchKMeans). The learned scaler and cluster centroids were then applied to both training and test samples within the fold, after which cluster-wise z-normalization was computed for all sensor channels to express deviations relative to the assigned regime. These steps yield a harmonized representation that supports the subsequent feature selection and HI construction.

4.3. Evaluation Protocol

All experiments followed a leave-one-cycle-out (LOCO) protocol at the cycle level. In each fold, a single cycle was held out for testing, and the remaining cycles were used for training. The preprocessing pipeline described in Section 4.2 was applied with all parameters estimated on the training partition of the fold. The learned scaler, cluster centroids, and regime statistics were then applied to the held-out test cycle.

Within each fold, feature selection (correlation ranking and derivative construction), health-indicator fitting (weighted Ridge), and monotonic calibration (isotonic mapping) were performed on the training partition only. The calibrated health indicator of each training cycle was used to construct the neighbor models for similarity-based estimation; predictions were subsequently generated along the tail of the held-out test cycle.

Performance was quantified using RMSE and MAE. Unless otherwise noted, the primary figure of merit is the average RMSE across LOCO folds, reported in the main comparison table; MAE is provided as a secondary indicator. All method comparisons (including the deep baseline) were conducted under the same preprocessing and evaluation protocol, enabling a direct assessment on a common footing. Hyperparameter settings referenced by the evaluation (e.g., number of neighbors, penalty/weighting, window lengths) are summarized in Section 4.4.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{RUL_i} - RUL_i\right)^2} \tag{17}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left|\,\widehat{RUL_i} - RUL_i\,\right| \tag{18}$$

4.4. Hyperparameters

The principal hyperparameters are grouped by pipeline stage and kept fixed across folds unless a one-factor sensitivity sweep is reported in the Results section.

Table 1. Default values used in experiments

Ridge regression regularization: α	10.0
Late-life weighting exponent: γ	3.0
Blend factor: λ	0.6
Smoothing window	9
Derivative window	5
Number of selected features	8
Number of neighbors	8
Distance exponent: p	2.0

4.5. Computational Environment

Experiments were conducted on a local workstation running macOS 15.6.1 (arm64) with an 8-core ARM processor and 18 GB of memory. All experiments were executed on CPU only, using Python and standard scientific libraries (NumPy, pandas, scikit-learn). The runtime of the leave-one-cycle-out evaluation was measured and the run time per fold was 1.0 second.

This efficiency highlights a key advantage of the similarity-based approach: competitive accuracy can be achieved without specialized hardware such as GPUs, enabling fast evaluation and deployment even in resource-constrained environments.

5. RESULTS

5.1. Main Comparison with Prior Work

Table 2 summarizes the average RMSE across LOCO folds for the F1 fault mode on the PHM2018 IME dataset under a common preprocessing and evaluation setup. The proposed similarity-based method achieved 453.53, which is substantially lower than the ATCN-LSTM baseline (597.35 s). Baseline values for RFR, MLP, LSTM, TCLSTM, DW-GRU, DW-GRU-FCs, HF-MS-MBTransformer, and ATCN-LSTM are taken from Darwish et al., 2024. The proposed result is obtained in this study. Among the methods listed in Table 2, the proposed approach attains the lowest error, providing a strong reference point for the subsequent analyses.

Table 2. Average RMSE for F1 fault mode

Model	RMSE (F1)
RFR	5476
MLP	5196
LSTM	1469
TCLSTM	601.47
DW-GRU	1014
DW-GRU-FCs	998
HF-MS-MBTransformer	646.42
ATCN-LSTM (baseline)	597.35
Proposed (Similarity-based)	**453.53**

5.2. Representative Predictions

Figure 2 shows representative prediction results for selected cycles. The trajectories show that the calibrated health indicator and neighbor-based estimation produce monotonic and smooth predictions that track degradation over the tail region. The predicted RUL trajectory closely follows the ground truth over the entire region, with only minor deviations observed.

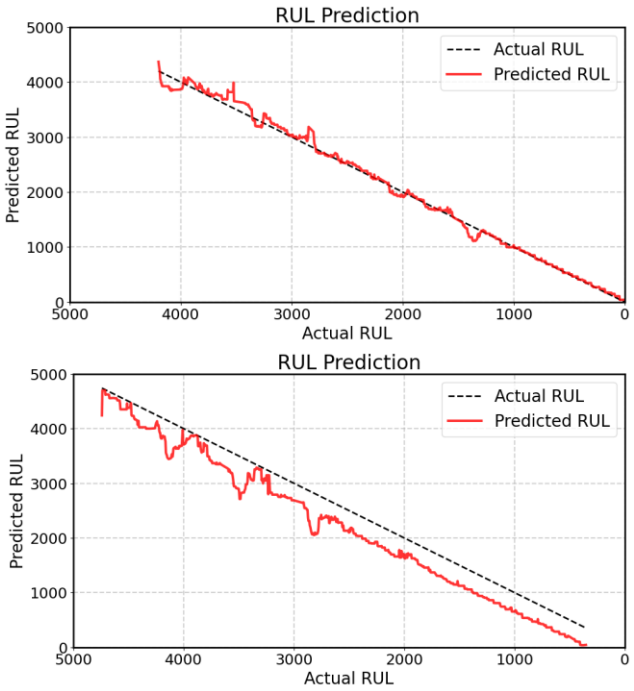
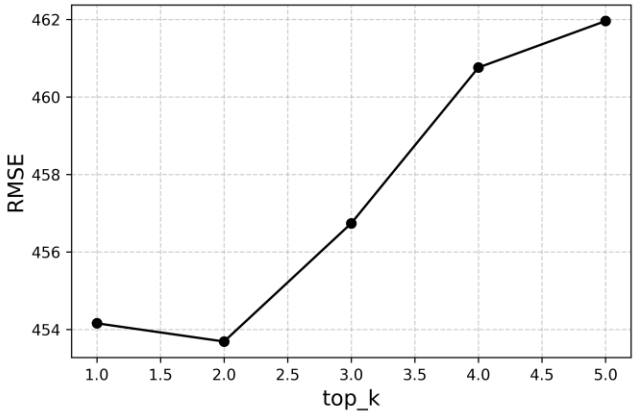


Figure 2. Examples of predicted RUL vs. ground truth for representative cycles

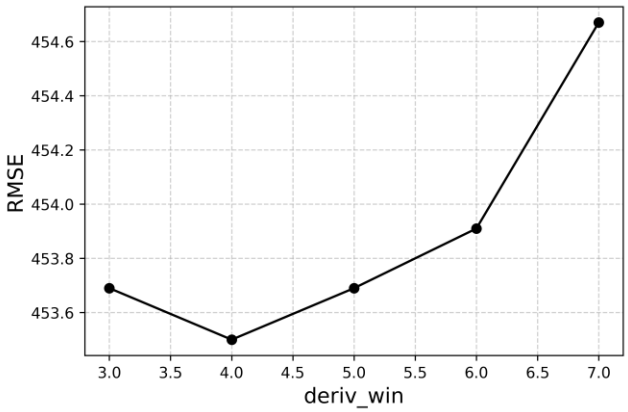
5.3. Sensitivity to Hyperparameters

One-factor-at-a-time sweeps were conducted around the default configuration in Table 1, varying a single parameter while holding the others fixed at the best setting identified in preliminary trials. Evaluation followed the LOCO protocol under the common preprocessing setup. Figure 3 reports RMSE as a function of each parameter: (a) number of selected features k, (b) derivative window, (c) smoothing window, (d) blend factor λ, and (e) late-life weighting exponent γ.

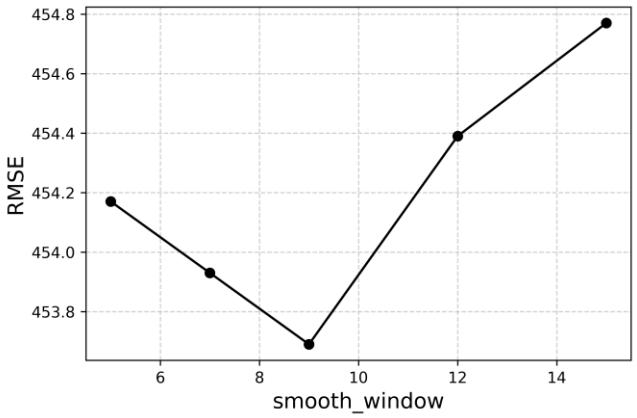
For the number of selected features k (Fig. 3a), performance improved as k decreased, with the best result achieved at k=1. This indicates that a single highly informative feature dominated the degradation representation for fault mode F1. The derivative window length (Fig. 3b) and the smoothing window length (Fig. 3c) produced moderate variations, with settings of 3–5 for the derivative window and 9 for the smoothing window providing the most favorable results. The blend factor λ (Fig. 3d) showed a tendency toward smaller values being more effective, while the late-life weighting exponent γ (Fig. 3e) had only minor influence within the tested range.



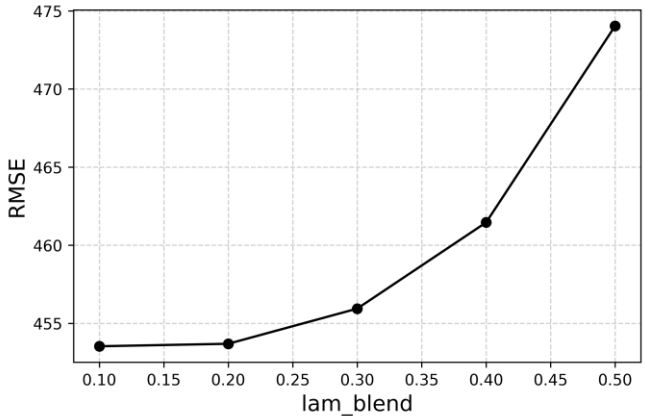
(a) Effect of the number of selected features k on RMSE.



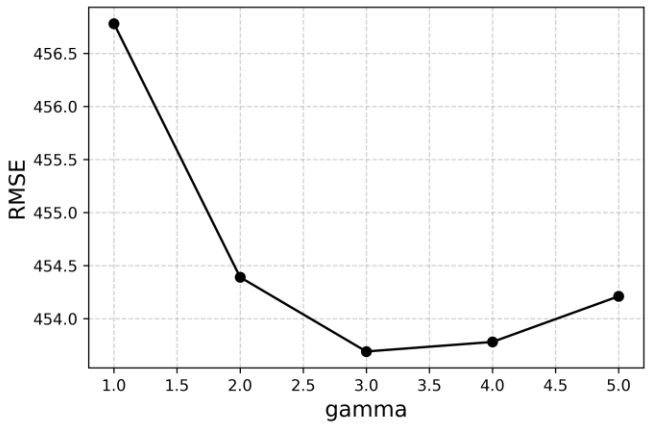
(b) Effect of derivative window length on RMSE.



(c) Effect of smoothing window length on RMSE.



(d) Effect of blend factor λ on RMSE.



(e) Effect of late-life weighting exponent γ on RMSE.

Figure 3. RMSE vs. hyperparameter settings. (a) number of selected features k, (b) derivative window length, (c) smoothing window length, (d) blend factor λ, and (e) weighting exponent γ

5.4. Summary of Findings

Across the PHM2018 IME evaluation under a common preprocessing and LOCO setup, the proposed similarity-based approach achieved lower average RMSE than the deep baseline used for reference (Table 2). Representative examples indicate that the calibrated, neighbor-based predictions are monotonic and smoothly track the degradation trajectory over the tail region (Figure 2). One-factor sensitivity analyses further show consistent trends: performance improves as the number of selected features k is reduced, with the best result obtained at k=2; shallow optima are observed for the derivative window in the 3–5 range and for the smoothing window around 9; smaller values of the blend factor λ are generally favorable within the tested range; and the late-life weighting exponent γ has only minor influence (Figure 3). Taken together, these results indicate that the method delivers competitive accuracy on F1 while maintaining simple, interpretable behavior of the predicted RUL trajectories.

6. DISCUSSION

The experimental results indicate that a similarity-based methodology can achieve competitive Remaining Useful Life prediction for semiconductor manufacturing equipment. On the PHM2018 IME task, the approach attained lower average RMSE than a strong deep baseline (ATCN-LSTM) under a common preprocessing and LOCO evaluation setup (Section 5.1). Representative trajectories further suggest that the calibrated, neighbor-based predictions evolve monotonically and smoothly over the tail region (Section 5.2), which is desirable for maintenance decision-making.

The reasons for the comparatively strong performance on F1 are not yet fully understood and merit further study. Potential contributing factors include differences in the degradation mechanisms specific to F1, the number and coverage of available cycles, and the signal-to-noise ratio of the most informative sensor channels. Future analyses could incorporate systematic feature-importance profiling, regime-specific trajectory comparisons, and statistical assessments of cycle-level variability to clarify whether the observed gains arise from intrinsic properties of the F1 failure mode or from characteristics of the data distribution (cf. Section 5.3).

The PHM2018 IME dataset comprises three failure modes (F1–F3). The present study concentrates on F1 to provide a clean and reproducible benchmark against a widely reported deep baseline (ATCN-LSTM) and to isolate the contributions of the similarity-based pipeline. This focus is consistent with prior practice in the IME literature when establishing method behavior under a single, well-defined condition. Generalization across modes remains to be investigated. Extension is expected to be straightforward in principle because the pipeline is mode-agnostic but will require mode-specific validation given differences in degradation mechanisms, class balance, and sensor salience.

The scope of the present study was limited to Tool 01_M02 and the F1 fault mode, following prior practice in the IME literature. While this setting provides a clear and reproducible benchmark for comparison with deep baselines, extending the applicability to other fault modes and tools remains an important direction. Promising avenues include data augmentation, few-shot learning strategies, and the use of pre-trained time-series representations to improve robustness when fewer cycles or more irregular signals are encountered.

An additional advantage lies in computational efficiency. The full LOCO evaluation (62 folds) completed in approximately one minute on a CPU-only workstation (≈1.0 s per fold), without specialized hardware (Section 4.5). This lightweight profile suggests that similarity-based prediction can be deployed rapidly and iterated frequently in production settings, providing a practical alternative to resource-intensive deep architectures while preserving interpretability of the prediction basis.

7. CONCLUSION

A similarity-based methodology for Remaining Useful Life prediction in semiconductor manufacturing equipment has been presented. The pipeline which consisting of regime-aware normalization, supervised health-indicator construction, monotonic calibration, and neighbor-based estimation was evaluated on the PHM2018 IME dataset (Tool 01_M02, F1) under LOCO evaluation. The approach achieved an average RMSE of 453.53, outperforming a strong deep baseline (ATCN-LSTM, 597.35) while preserving interpretability of the prediction basis and a simple implementation pathway.

Beyond accuracy, the method exhibits a favorable computational profile: the full 62-fold evaluation completed in approximately one minute on a CPU-only workstation (≈1.0 s per fold), suggesting that rapid iteration and deployment are feasible on commodity hardware. These characteristics position similarity-based prediction as a lightweight and industrially practical alternative to resource-intensive deep architectures for complex manufacturing systems.

The present study focused on F1 to provide a clear and reproducible point of reference. Extending the methodology and evaluation to F2 and F3, as well as to additional tools and operating conditions, represents a natural next step to assess cross-mode generalization. Promising directions include data augmentation, few-shot learning strategies, and the use of pre-trained time-series representations, as well as hybrid frameworks that combine the transparency of similarity-based estimation with the representational capacity of modern deep models.

REFERENCES

Berghout, T., & Benbouzid, M. (2022). A systematic guide for predicting remaining useful life with machine learning. Electronics, 11(7), 1125.

Darwish, A. (2024). A data-driven deep learning approach for remaining useful life in the ion mill etching process. Sustainable machine intelligence journal, 8(2), 14-34.

Ferreira, C., & Gonçalves, G. (2022). Remaining Useful Life prediction and challenges: A literature review on the use of Machine Learning Methods. Journal of Manufacturing Systems, 63(May), 550-562.

Hsu, C. Y., Lu, Y. W., & Yan, J. H. (2022). Temporal convolution-based long-short term memory network with attention mechanism for remaining useful life prediction. IEEE Transactions on Semiconductor Manufacturing, 35(2), 220-228.

Liu, C., Zhang, L., Li, J., Zheng, J., & Wu, C. (2021). Two-stage transfer learning for fault prognosis of ion mill etching process. IEEE Transactions on Semiconductor Manufacturing, 34(2), 185-193.

Liu, Y., Wen, J., & Wang, G. (2025). A comprehensive overview of remaining useful life prediction: From

traditional literature review to scientometric analysis. Machine Learning with Applications, 100704.

Xue, B., Xu, H., Huang, X., Zhu, K., Xu, Z., & Pei, H. (2022). Similarity-based prediction method for machinery remaining useful life: A review. The international journal of advanced manufacturing technology, 121(3), 1501-1531.

Yuan, Z., & Wang, R. (2023, August). A Squeeze-and-Excitation and Transformer Based Model for Remaining Useful Life Prediction in Ion Mill Etching Process. In 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE) (pp. 1-6). IEEE.

Yuan, Z., & Wang, R. (2024). Multi-scale and multi-branch transformer network for remaining useful life prediction in ion mill etching process. IEEE Transactions on Semiconductor Manufacturing, 37(1), 67-75.

Wang, T., Yu, J., Siegel, D., & Lee, J. (2008, October). A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In 2008 international conference on prognostics and health management (pp. 1-6). IEEE.

Wang, T. (2010). Trajectory similarity-based prediction for remaining useful life estimation. University of Cincinnati

Wu, F., Wu, Q., Tan, Y., & Xu, X. (2024). Remaining useful life prediction based on deep learning: a survey. Sensors, 24(11), 3454.

Wu, S., Jiang, Y., Luo, H., & Yin, S. (2021). Remaining useful life prediction for ion etching machine cooling system using deep recurrent neural network-based approaches. Control Engineering Practice, 109, 104748.

BIOGRAPHIES

Takanobu Minami received his B.S. and M.S. degrees in mechanical engineering from Kyoto University in 2008 and in 2011, respectively. Currently, he is pursuing his Ph.D. degree in mechanical engineering with the University of Maryland, College Park, MD, USA, and is employed as an engineer in Komatsu Ltd. His research interests include machine learning, deep learning, prognostics and health management, and industrial AI.

Jay Lee is Clark Distinguished Professor and Director of the Industrial AI Center in the Mechanical Engineering Dept. of the Univ. of Maryland College Park. His research is focused on intelligent analytics of complex systems including highly-connected industrial systems including energy, manufacturing, healthcare/medical, etc. He has been working with medical school in Traumatic Brain Injury (TBI) using multi-dimension data for predictive assessment of patient in ICU with funding from NIH and NSF. Previously, he served as an Ohio Eminent Scholar, L.W. Scott Alter Chair and Univ. Distinguished Professor at Univ. of Cincinnati. He was Founding Director of National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC) on Intelligent Maintenance Systems during 2001-2019. IMS Center pioneered industrial Ai-augmented prognostics technologies for highly-connected industrial systems and has developed research memberships with over 100 global company since 2000 and was selected as the most economically impactful I/UCRC in the NSF Economic Impact Study Report in 2012. He is also the Founding Director of Industrial AI Center. He mentored his students and developed a number of start-up companies including Predictronics through NSF iCorp in 2013 and has won 1st Place for PHM Society Data Challenges competition 5 times He was on leave from UC to serve as Vice Chairman and Board Member for Foxconn Technology Group (ranked 26th in Global Fortune 500) during 2019-2021 to lead the development of Foxconn Wisconsin Science Park (~$1B investment) in Mt. Pleasant, WI. In addition, he advised Foxconn business units to successfully receive five WEF Lighthouse Factory Awards since 2019. He is a member of Global Future Council on Advanced Manufacturing and Production of the World Economics Council (WEF), a member of Board of Governors of the Manufacturing Executive Leadership Council of National Association of Manufacturers (NAM), Board of Trustees of MTConnect, as well as a senior advisor to McKinsey. Previously, he served as Director for Product Development and Manufacturing at United Technologies Research Center (now Raytheon Technologies Research Center) as well as Program Director for a number of programs at NSF. He was selected as 30 Visionaries in Smart Manufacturing in by SME in Jan. 2016 and 20 most influential professors in Smart Manufacturing in June 2020, SME Eli Whitney Productivity Award and SME/NAMRC S.M. Wu Research Implementation Award in 2022