PHM-AP 2025 Data Challenge Technical Report: Cumulative Incremental Machine Learning Guided Flank Wear Prediction

Jonathan Adam Rico[1], Nagarajan Raghavan[2], and Senthilnath Jayavelu[3]

[1,2] *Engineering Product Development, Singapore University of Technology and Design, Singapore*
sturicoja@a-star.edu.sg
nagarajan@sutd.edu.sg

[1,3] *Institute for Infocomm Research, Agency for Science, Technology, and Research, Singapore*
j_senthilnath@a-star.edu.sg

1. Introduction

The PHM-AP 2025 data challenge is about predicting the flank wear of the cutting tool of a CNC machine using time series signals from accelerometer and acoustic emission (AE) sensors. This data challenge has several bottlenecks such as *limited labeled data* where the cumulative flank wear labels are only available for specific cuts: 1, 6, 11, 16, 21, and 26. Second, there is *limited samples* with only 6 cutting tools in the training data and 3 cutting tools in the evaluation data. Third, there are *limited sensor signals* with only three-axis accelerometer and acoustic emission signals and no signals from force or thermal sensors. Finally, solutions are bounded by the *conditions* of the competition rules such as 1-hour maximum Docker run, 6 GB memory, CPU only, and no internet access.

This technical report provides a summary of our proposed solution, Cumulative Incremental Machine Learning (CIML), to the PHM-AP 2025 data challenge. Our solution achieved 11.461 root-mean squared error (RMSE), 9.508 mean absolute percentage error (MAPE), and coefficient of determination (R^2) $= 0.875$ in the final evaluation securing a tie for the 3rd-place award.

2. Methodology and Modeling Approach

We propose the CIML framework for flank wear prediction. This framework utilizes both cumulative (summation of flank wear at every five cuts) and incremental cut for flank wear prediction. The cumulative approach has real labels but limited to 30 training samples. In the incremental approach, we expand the data from 30 to 150 training samples but with added uncertainty due the pseudo labels. The pseudo labels are generated using a weak model trained only on the first cuts and sum scaled to match the available cumulative wear data. Then we perform data preprocessing with denoising the signals using bandpass filter; extracting statistical, spectral, domain-specific, and controller features further discussed in Section 3. Finally, we used linear and non-linear machine learning model for flank wear prediction with 6-fold cross validation where each cutting tool is a fold, further discussed in Section 4.

3. Feature Engineering

3.1. Data Preprocessing

The acoustic emission signals are already envelope extracted signals while the three-axis accelerometer signals are raw data. Even so, performing envelope extraction in three-axis accelerometer signals worsen the performance. Thus, only bandpass filtering was done for all the signals. The acoustic emission signals are dominated by low frequency sounds yet we are more interested in the high frequency bursts. Thus, we performed Butterworth high-pass filter with cut-off frequency of 5 kHz. For the accelerometer signal, we used Butterworth bandpass filter from 50 Hz to 7 kHz.

3.2. Feature Extraction

We extracted 82 features from sensor and controller data. These features are normalized by taking their relative difference with the corresponding tool's first cut's features. Both the raw and normalized features are considered, yielding a total of 164 features.

Sensor Features. We extracted 126 (raw and normalized) sensor features which include statistical features such as mean, maximum, minimum, standard deviation, variance, root mean square (RMS), peak-to-peak values, skew, kurtosis, crest factor, entropy; frequency-domain features such as the spectral centroid, spectral bandwidth, spectral entropy, and

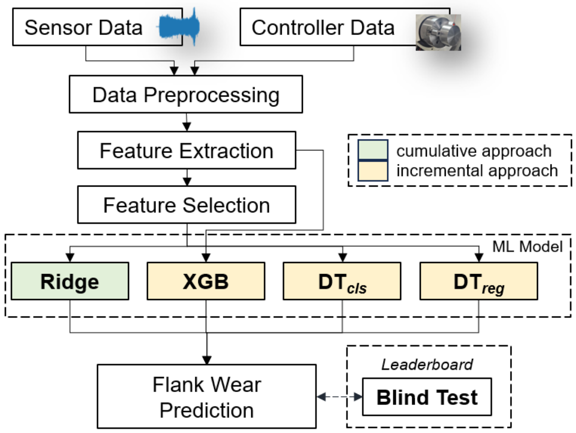

Figure 1. CIML for flank wear prediction.

dominant frequency from accelerometer XYZ and AE signals; and AE-specific features such as median, peak count, and signal energy.

Controller Features. We extracted 38 (raw and normalized) controller features which include the mean, maximum, and sum of X load, Y load, Z load, feedrate, main spindle speed, and main spindle load of each cut; and effective cutting time.

3.3. Feature Selection

All features were utilized in XGBoost while feature selection is performed for Ridge and Decision Tree models. Five (5) features were selected for Ridge Regression by random search of feature combinations that minimize the hold-out error, these include 'AE skew t2', 'AE entropy t2', 'AE entropy t3', 'differenced accelerometer X mean t1', and 'accelerometer X mean t4' from features with top correlation with the target cumulative wear. The suffix "_tn" indicates the nth signal in groups of 5 consecutive cuts. For Decision Tree models, we empirically selected two (2) features, accelerometer RMS in X and Y axes, which are sufficient to detect and predict the flank wear of weak cutting signals.

4. MODEL DEVELOPMENT

The flank prediction is performed using an ensemble of four models: Ridge Regression, Extreme Gradient Boosting (XGB), Decision Tree Classifier (DT$_{cls}$) and Regressor (DT$_{reg}$).

4.1. Training Methodology

We trained four models: XGB for incremental flank wear prediction, Ridge for cumulative wear prediction, and Decision Trees for predicting incremental flank wear from weak cutting signals. We used Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) to interpolate monotonic incremental wear predictions from the cumulative wear predictions. We performed ensembling for the four models using the average of XGB and Ridge while the Decision Trees handle the edge cases for weak cutting signal detection (classification) and prediction (regression). Figure 1 shows the overall pipeline for the CIML flank wear prediction. All training are optimized using RMSE loss function. Hyperparameters are tuned using Optuna for XGB, RidgeCV for Ridge model, and no tuning for Decision Tree models. XGB is trained with early stopping criteria to save the best model based on out-of-fold validation score. Local validation scheme uses 6-fold cross validation on the cumulative flank wear.

4.2. Evaluation

As provided by the PHM-AP committee, we used the official metrics RMSE, MAPE, and R^2 in the data challenge for evaluation. Table 1 shows the evaluation performance of our single and ensemble models.

Table 1. Evaluation performance of cumulative approach (CA) and incremental approach (IA).

Approach/Model	RMSE	MAPE	R^2
CA (Ridge)	17.166	16.859	0.720
IA (XGB)	15.621	13.385	0.768
CA + IA (Ridge + XGB)	13.098	12.393	0.837
CIML	**11.461**	**9.508**	**0.875**

5. KEY INSIGHTS

We noticed that the controller data are not aligned with the sensor data in timestamps, thus, it is essential to detect the regions using signal RMS instead of relying on the controller data timestamps. Our modification of the data loader also speeds up the inference by eight fold. Noisy and invalid sensor data are found in accelerometer Z-axis of trainset 4 for cuts 17-22 and thus removed from the training pipeline. The bandpass filtering especially made an impact on acoustic emission where the low frequency signals are removed. In this way, the model focuses on capturing the sudden high-frequency bursts in the acoustic emission signal. Table 1 shows the improvement from cumulative approach (CA) 17.166 RMSE and incremental approach (IA) 15.621 RMSE to (CA + IA) 13.098 RMSE. The CA is limited to 30 samples, making training highly susceptible to overfitting, while the IA with pseudo targets adds more data but have high uncertainty. Thus, by combining both approach, we develop a more generalized solution. Furthermore, incorporating outlier detection and prediction of weak cutting signals improved the performance from 13.098 RMSE (Ridge + XGB) to 11.461 RMSE (CIML). Finally, our work is limited to machine learning models since our best deep learning model only achieved 18 RMSE and more complex models do not satisfy the memory limitations of the competition.