Ensemble Machine Learning Algorithms for Remaining Useful Life Prediction

Richard Duffy[1,2,*], **Sagar Nikam**[2]
[1]North West Regional College, Strand Road, Londonderry BT48 7AL, UK
[2]School of Computing, Engineering and Intelligent Systems, Ulster University, Derry/Londonderry, BT487JL, UK
* Correspondence: richard.duffy@nwrc.ac.uk

ABSTRACT

Accurate cutting-tool degradation prediction is essential for predictive maintenance in Industry 4.0. This paper presents an explainable ensemble framework for direct Remaining Useful Life (RUL) prediction using the NASA Milling Dataset. Wavelet Packet Decomposition (WPD), training-only feature selection and grouped cross-validation were combined to support robust model evaluation. Extra Trees, Gradient Boosting, XGBoost, LightGBM and CatBoost were evaluated alongside ensemble methods. The stacking ensemble achieved the best performance ($R^2 = 0.898$, mean absolute error (MAE) = 3.790 min, root mean square error (RMSE) = 5.443 min). SHapley Additive exPlanations (SHAP) analysis identified influential machining and wavelet-derived sensor features, improving model interpretability.

KEYWORDS: Predictive Maintenance; Remaining Useful Life; Ensemble Learning; Explainable AI; Wavelet Packet Decomposition.

1. INTRODUCTION

Cutting-tool wear affects machining quality, production efficiency and maintenance costs. Conventional maintenance can result in premature tool replacement or excessive wear, increasing downtime and operating costs [1], [2]. Machine learning enables tool degradation to be estimated from sensor measurements, while ensemble learning combines complementary models to improve prediction robustness and generalisation.

This study uses the NASA Milling Dataset [3] to develop an explainable ensemble framework for direct RUL prediction. Extra Trees, Gradient Boosting, XGBoost [4], CatBoost [5] and ensemble methods are compared, while SHAP analysis [6] identifies the most influential wavelet-derived sensor features and improves model interpretability.

2. MATERIALS AND METHODS

Multi-sensor signals from the NASA Milling Dataset [3] were standardised before level-3 Daubechies-4 WPD extracted statistical and time-frequency features. Correlation analysis supported feature assessment, while training-only Extra Trees selection retained up to 100 features per GroupKFold partition. Extra Trees, Gradient Boosting, XGBoost [4], LightGBM and CatBoost [5] were evaluated using grouped cross-validation. Their Out-of-Fold (OOF) predictions were combined using an ElasticNet meta-learner ($\alpha = 0.001$, l1_ratio = 0.5), alongside weighted and voting ensembles. SHAP analysis [6] was then used to interpret RUL predictions.

3. RESULTS

The OOF stacking ensemble achieved the best RUL performance (MAE = 3.790 min, RMSE = 5.443 min, $R^2 = 0.898$), followed by the weighted ensemble ($R^2 = 0.876$) and Extra Trees ($R^2 = 0.874$). SHAP ranked workpiece material, depth of cut (DOC), machining time and spindle acoustic-emission (AE) wavelet energy ratio highest, with mean |SHAP| values of 0.1968, 0.0940, 0.0596 and 0.0593 min, respectively. These values represent each feature's contribution to predicted RUL in minutes, not the physical units of the input features. Table 1 compares model performance, while Fig. 1 shows the principal SHAP contributions. GroupKFold maintained independence between machining cases.

© The Author(s) 2026
D. Nikam et al. (eds.), *Proceedings of the Global Symposium on Adaptive Manufacturing (GSAM 2026)*, Advances in Engineering Research 308,
https://doi.org/10.2991/978-94-6239-813-9_21

Table 1. Performance comparison of the evaluated machine learning algorithms for direct RUL prediction.

Algorithm (Python Library)	MAE	RMSE	R^2
OOF Stacking Ensemble (scikit-learn)	3.790	5.443	0.898
Weighted Ensemble (NumPy / scikit-learn)	4.050	5.996	0.876
Extra Trees Regressor (scikit-learn)	4.142	6.044	0.874
CatBoost Regressor (catboost)	4.201	6.208	0.867
Voting Ensemble (scikit-learn)	4.591	6.879	0.837
XGBoost Regressor (xgboost)	5.396	8.199	0.769
LightGBM Regressor (lightgbm)	5.178	8.213	0.768
Gradient Boosting Regressor (scikit-learn)	5.782	8.606	0.745

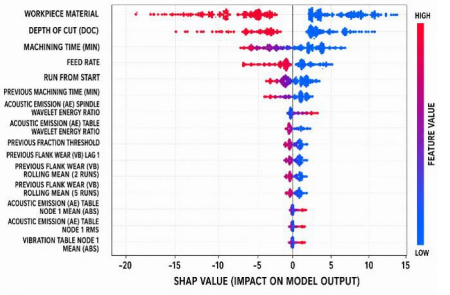


Fig. 1. SHAP bee swarm plot illustrating the contribution of the most influential variables to RUL prediction.

4. DISCUSSION

The stacking ensemble combined Extra Trees, Gradient Boosting, XGBoost, LightGBM and CatBoost using an ElasticNet meta-learner, improving R^2 from 0.874 to 0.898 and reducing MAE from 4.142 to 3.790 min. Base-model predictions were generated out-of-fold using GroupKFold, reducing leakage between machining cases and supporting robust evaluation. The improved accuracy demonstrates the potential of ensemble learning for direct RUL prediction, providing useful information for tool-change scheduling and condition-based maintenance. SHAP analysis identified influential machining and wavelet-derived sensor features. However, validation was limited to the NASA benchmark dataset, and further nested and independent industrial validation is required to confirm generalisation.

5. CONCLUSION

The proposed explainable RUL framework combined level-3 Daubechies-4 WPD, training-only feature selection, grouped validation and ensemble learning using five base learners and an ElasticNet meta-learner. The stacking ensemble achieved R^2 = 0.898, MAE = 3.790 min and RMSE = 5.443 min, outperforming the evaluated individual and alternative ensemble models. SHAP identified physically relevant machining, progression and sensor-derived predictors, while direct RUL prediction in minutes provides an interpretable basis for maintenance scheduling. Future work should validate the complete stacking framework using nested cross-validation and independent industrial machining data and investigate transformer-based temporal models.

REFERENCES

[1] Carvalho, T.P.; Soares, F.A.A.M.N.; Vita, R.; Francisco, R.P.; Basto, J.P.; Alcalá, S.G.S. A systematic literature review of machine learning methods applied to predictive maintenance. Comput. Ind. Eng. 2019, 137, 106024. doi: 10.1016/j.cie.2019.106024.

[2] Munaro, R.; Attanasio, A.; Del Prete, A. Tool Wear Monitoring with Artificial Intelligence Methods: A Review. J. Manuf. Mater. Process. 2023, 7, 129. doi: 10.3390/jmmp7040129.

[3] Agogino, A.; Goebel, K. Milling Data Set. BEST Lab, University of California, Berkeley; NASA Prognostics Data Repository, NASA Ames Research Center: Moffett Field, CA, USA, 2007.

[4] Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. doi: 10.1145/2939672.2939785.

[5] Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. In Advances in Neural Information Processing Systems 31, Montréal, Canada, 3–8 December 2018; pp. 6638–6648.

[6] Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

