A Feature-Engineering-Based Machine Learning Approach for Cutter Flank Wear Prediction under Data-Scarce Conditions

Paula Mielgo[1], Marcos Quinones-Grueiro[2], Anibal Bregon[3], Austin Coursey[4], Carlos J. Alonso-González[5] and Gautam Biswas[6]

[1,3,5] *Department of Computer Science, University of Valladolid, Valladolid, Spain*
paula.mielgo@uva.es, anibal.bregon@uva.es, calonso@uva.es

[2,4,6] *Institute for Software Integrated Systems, Vanderbilt University, Nashville, USA*
marcos.quinones.grueiro@vanderbilt.edu,austin.c.coursey@vanderbilt.edu,gautam.biswas@vanderbilt.edu

ABSTRACT

Accurate estimation of tool wear in machining processes is essential to ensure product quality and optimize maintenance strategies. This work presents a Machine Learning methodology for the PHM-AP 2025 Data Challenge. The objective of the challenge is the cutter flank wear prediction in a CNC mill-turn machine using accelerometer, acoustic emission, and controller data. The training data consists of six datasets with a limited number of labeled samples, resulting in a few-shot learning scenario. To address these constraints, a manual feature extraction method is proposed. Features are computed by aggregating data from the controller and sensors in the time and frequency domains across five-cut intervals. In this way, the wear behavior is captured, and the sensitivity to missing data is reduced. Then, an optimization process is performed to select the most relevant features based on correlation values. These 14 identified features are used to fit a Multilayer Perceptron through a leave-one-dataset-out cross-validation process. Results reveal variability between training sets, with pronounced errors in the 17-21 cutting interval in four datasets. However, in the evaluation stage, the model achieved a competitive performance: RMSE of 11.486, MAPE of 8.518, and R^2 of 0.875, placing fourth in the challenge.

1. INTRODUCTION

Wear prediction represents a fundamental challenge in modern manufacturing. Accurately determining the wear level of a production tool or component in production time is essential not only to maximize its useful lifetime but also to ensure the quality of manufactured products (Zhou, Liu, Yu, Liu, & Quan, 2022). In Computer Numerical Control (CNC) mill-turn machines, a number of cycles or processed parts were traditionally set to replace the cutting tool. However, the actual wear of a cutting tool is not uniform over its useful lifetime (Colantonio, Equeter, Dehombreux, & Ducobu, 2021), presenting variability caused by external factors such as environmental dust, operating conditions, or even structural or quality differences in the tool. Hence, systematically increasing this fixed value to maximize component lifetime without any criteria entails a risk of catastrophic failures with higher operational costs than those associated with early replacement (Zhou & Xue, 2018). As a consequence, the effective prediction of wear level in future production cycles is crucial for making optimal decisions about the best time to replace the tool, enabling the development of predictive maintenance strategies rather than scheduled ones (Traini, Bruno, & Lombardi, 2021).

Wear prediction has traditionally been addressed through empirical approaches, fixed rules, and physics-based models that incorporate variables such as pressure, temperature, cutting forces, or machining dynamics (Usui, Shirakashi, & Kitagawa, 1984) (Koren, Ko, Ulsoy, & Danai, 1991). With the rise of sensor technology, devices such as accelerometers, dynamometers, load cells, and acoustic emission sensors began to be used to capture data from CNC machines and to develop Machine Learning (ML) models, where techniques such as Support Vector Machines, Random Forest, and Artificial Neural Networks enhanced predictive capability (Jones & Cao, 2025) (Wu, Jennings, Terpenny, Gao, & Kumara, 2017). Recently, the advent of Deep Learning (DL) has enabled models to learn from data without manual feature extraction, using Convolutional Neural Networks, Long-Short Term Memory models, or Transformers to achieve more accurate results (Xu, Wang, Zhong, Ming, & Chen, 2021) (Hirsch & Friedrich, 2024). Nevertheless, their effectiveness is often limited when large datasets are not available. Finally, hybrid models have also been used, combining physical models with ML/DL architectures to improve generalization and increase robustness under varying conditions (Yang, Pattipati, Awasthi, & Bollas, 2022) (Y. Zhang & Zhu, 2022).

In this paper, we propose a methodology specifically developed for the PHM-AP 2025 Data Challenge. Unlike approaches based on deep architectures, whose effectiveness is hindered by the availability of only six complete trajectories and a very limited number of labels per tool, our method relies on feature extraction from acceleration signals, acoustic emission data, and CNC controller variables, combining descriptors from the temporal, frequency, and statistical domains. These features are aggregated over five-cut intervals to capture the wear evolution and mitigate the impact of missing data, transforming the high-frequency series into a compact tabular representation. A systematic feature selection process is then applied, followed by training an MLP model using leave-one-dataset-out cross-validation. This approach achieves a balance between predictive accuracy and overfitting control, ranking fourth in the competition and demonstrating that, in data-scarce scenarios, feature-engineering-based strategies coupled with simple models can outperform more complex alternatives.

The remainder of the paper is organized as follows. Section 2 introduces the PHM-AP 2025 Data Challenge and its evaluation setting. Section 3 provides a detailed description of the datasets, including controller signals, sensor measurements, and wear labels. Section 4 presents the proposed methodology, covering feature extraction, feature selection, and model optimization. Section 5 details the experimental results achieved on the training and evaluation data. Finally, Section 6 presents the main conclusions and outlines future directions for this work.

2. DATA CHALLENGE DESCRIPTION

The PHM-AP 2025 Data Challenge addresses the problem of predicting cutter flank wear in machining operations. The dataset was collected during cutting tests conducted on a DMG Mori NTX2500, where each insert performed 26 consecutive cuts on stainless steel workpieces. For every cut, signals from a triaxial accelerometer and an acoustic emission sensor are recorded at 25.6 kHz. Furthermore, data from the CNC controller, including loads, speeds, and CNC states, is also captured. In those experiments, wear is physically measured after specific cuts (1, 6, 11, 16, 21, and 26), providing temporally spaced labels. The objective of the challenge is to develop models to predict the complete wear trajectory from these measurements, together with a single initial wear label (cut 1).

For the training stage, six datasets are available, providing sensor and controller data across all 26 cuts, along with their wear labels. On the other hand, the evaluation stage is performed on three datasets containing only data from the first cut, with the remaining information hidden. The evaluation procedure was performed in an environment with limited computational resources. Each team was allowed to make two submissions per day by providing a Docker image that was executed on a CPU-only system with limited memory, with a maximum time limit for completing the inference process. Submissions were required to load the controller and sensor data and generate predictions for cuts 2-26 in each evaluation set. The performance was evaluated using RMSE, MAPE, and R^2. These constraints encourage the development of efficient and robust methods for tool wear prognosis in real industrial environments.

3. DATA DESCRIPTION

This section presents an analysis of the data provided in the challenge, including controller and sensor data, as well as the target variable. As mentioned earlier, each experiment consist of 26 cuts. Each cut is composed of a sequence of steps, during which several measurements are taken. Each type of data is captured at a different granularity; an overview is shown in Figure 1.

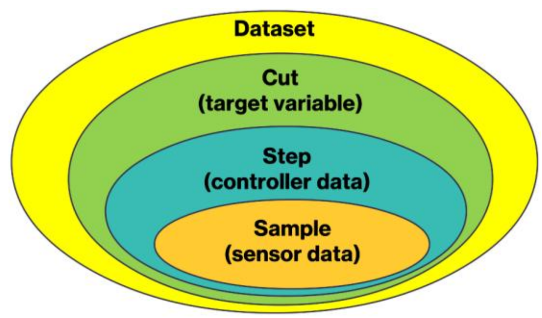


Figure 1. Data granularity

3.1. Controller Data

Controller data is recorded at every step of each cut. It consists of machine controller records from the machining process, including information on program status, operating modes, and spindle speeds and loads. It also contains timestamps recording the start and end of each cut and each step, allowing the sensor signals to be linked to the actual machining conditions at each moment.

Among the variables it contains, there are many that are constant, such as those related to status, program name, or operating mode, and were therefore discarded. The variables retained include the feed rate, the main spindle speed, and loads on the X, Y, Z, and B axes. In terms of data completeness, the step records are not always consistent across cuts. Specifically, information for a single step is missing in 19 cuts, while information for two steps is missing in another 6 cuts.

3.2. Sensor Data

As mentioned before, sensor data is recorded at a sampling rate of 25,600 Hz across four channels: a tri-axial accelerometer (X, Y, and Z) and an acoustic emission sensor. The X axis represents the neutral axis, Y is the feed axis, and Z corresponds to the cutting axis. A timestamp is also recorded for each measurement. Missing values appear in two different forms. On the one hand, there are single missing records within step intervals. These occur in all datasets and represent 16.82% of the total records; however, as they appear as isolated points, they do not affect the characterization of the overall signal behavior. On the other hand, dataset 6 lacks information for cuts 1, 2, 7, 12, 17, and 22; specifically, between 74% and 80% of the data are missing in each case. This issue must be carefully addressed in the proposed methodology, as similar situations may occur in the evaluation data.

3.3. Target

The target variable is the flank wear of the cutting tool. As mentioned above, measurements are captured after cuts 1, 6, 11, 16, 21, and 26. Figure 2 shows the trend in the cumulative values across the six training datasets. No clearly separable degradation modes are observed across the six trajectories, as their wear curves follow comparable monotonic patterns.

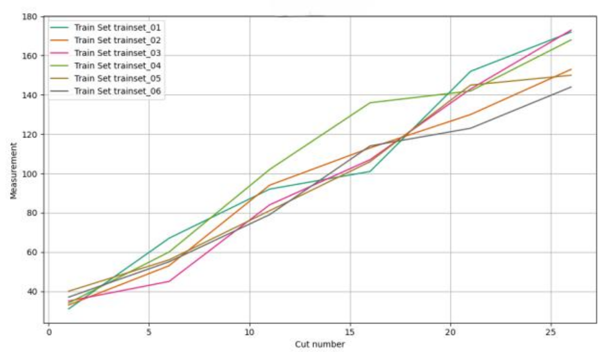


Figure 2. Flank wear for training data

4. PROPOSED METHOD AND EXPERIMENTAL PROCEDURE

Recent work on tool wear prediction in machining has explored a wide range of approaches, from signal processing combined with classical ML to DL architectures. (C. Zhang, Yao, Zhang, & Jin, 2016) process triaxial accelerometer signals using wavelet analysis to extract discriminative features, which are then fed into a Neuro-Fuzzy Network to estimate wear and remaining useful life (RUL). Other studies, often built upon the PHM 2010 Data Challenge dataset (Li, 2021), combine dynamometer, accelerometer, and acoustic emission signals to train DL models for RUL prediction. (Si, Mu, & Si, 2024) transform sensor signals into power spectral density (PSD) maps and then apply hybrid deep architectures, such as CNN and ViT combination, to simultaneously capture local and global features, while (Martínez-Arellano, Terrazas, & Ratchev, 2019) convert 1D signals into image representation using Gramian Angular Summation Fields for subsequent CNN classification.

However, these methodologies cannot be directly applied to the problem presented in PHM-AP 2025 due to its few-shot nature, as only six complete datasets are available for training and a very limited number of labels per tool. This significantly limits the suitability of deep models and increases the risk of overfitting, particularly when learning from high-frequency time-series data. To address these constraints, our approach relies on manual feature extraction from sensor data in both time and frequency domains, along with the controller data. Once these features are computed, a selection process of the most relevant ones is performed, and an MLP model is trained[1] to predict wear for cuts 6, 11, 16, 21, and 26. Partial wear per cut was used as the target instead of cumulative wear, as this is the variable required for the final evaluation. Each stage of the process is described in detail below.

4.1. Feature Extraction

Since the wear value at each labeled point includes the cumulative degradation of the five immediately preceding cuts, feature extraction was performed by aggregating data across these five-cut intervals. Specifically, features summarizing intervals corresponding to the sequences of cuts 2–6, 7–11, 12–16, 17–21, and 22–26 were constructed. This approach offers two main advantages. First, it allows the data to be transformed into a tabular format, thereby avoiding the need for sequential or time-series-based models. Second, it reduces the impact of missing values because, even in the presence of cuts with high percentages of missing records (as in training set six), aggregation into wide intervals captures the overall trend in signal behavior. As a consequence, no explicit imputation or removal of intervals was performed; features were computed using only the available measurements within each five-cut interval.

Features were extracted from three sources: temporal-domain sensor data, frequency-domain sensor data, and controller data. For each of the four sensor channels (acceleration in X, Y, and Z, and acoustic emission), 20 temporal-domain features were generated from descriptive statistics, shape measures, energy, entropy, and metrics related to signal dynamics. In the frequency domain, the signals were first transformed using the Fast Fourier Transform (FFT). Then, 15 features were computed, including basic spectral magnitudes and advanced descriptors such as spectral centroid, dispersion, or spectral flatness. Additionally, for each CNC controller variable, 23 statistical and dynamic characteristics were extracted, including measures of variability, temporal gradients, and nonlinear metrics.

1. All experiments were performed on an Intel Xeon Silver 4310 (2.10 GHz) with 32 GB of RAM, and an NVIDIA A40 (48 GB) was used to accelerate training. The software environment was based on Python 3.9-slim, with PyTorch 2.5.1 and Scikit-Learn 1.6.1 as the main frameworks.

Each five-cut interval is therefore represented by a tabular feature vector of 301 elements, providing a compact but informative characterization of tool behavior. For brevity, not all extracted features are described. However, the variables selected for model training are detailed in the following subsection.

4.2. Feature Selection

Due to the high dimensionality of the extracted features and the limited number of labeled samples, a feature selection process was applied. First, the correlation between each feature and the target variable was computed. Then, an iterative procedure was performed, starting with the two features showing the highest correlation, the next most correlated feature was added at each step. To determine the optimal number of features, cross-validation was conducted using a simple Multilayer Perceptron (MLP) topology under a leave-one-dataset-out scheme, in which one of the six datasets was reserved for evaluation at each iteration. The challenge metrics (RMSE, MAPE, and R^2) were calculated at each iteration. Table 1 presents the top five results obtained during the feature selection process, according to the challenge ranking criteria. The model using 14 features achieves the best overall performance. These features, described in Table 2, capture phenomena including force irregularity (DCOUNT, XP2P), increased cutting effort (XMAX), consistent changes over time (DMEAN), shifts in vibration energy (SMED), or the appearance of impulsive or unstable events (XCREST, XCLEAR, XKURT, XRMS, XSTD).

Table 1. Feature selection results.

Feature count	Mean RMSE	Mean MAPE	Mean R2
14	5.984	28.831	0.600
15	6.300	30.661	0.571
16	6.661	30.509	0.516
4	7.019	28.516	0.464
19	6.817	32.205	0.496

4.3. Model Optimization

The last step was to train a model using these 14 selected features. An MLP was optimized with Optuna, using cross-validation to evaluate and select the best hyperparameter configuration. The optimized parameters included the number of layers, hidden layer size, activation function, learning rate, weight decay, and batch size. An early-stopping mechanism based on the validation loss was used to prevent overfitting during training. During this process, features corresponding to training set 4 in the 17–21 interval consistently degraded model performance; subsequent inspection of these feature values confirmed this anomalous behavior. For this reason, this interval was treated as an outlier and excluded from model training. Final results are shown in Table 3.

In addition, other classic ML models were evaluated, such as Support Vector Machine (SVM), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), and Random Forest (RF). In all cases, the performance was lower than that obtained with the MLP as reported in Table 4.

Table 4. ML models results.

Model	Mean RMSE	Mean MAPE	Mean R^2
MLP	6.027	28.063	0.612
SVM	7.486	35.838	0.422
XGBoost	8.440	41.554	0.293
LightGBM	9.074	43.134	0.183
CatBoost	9.189	45.341	0.170
RF	9.439	47.347	0.093

Therefore, the MLP architecture and its optimal hyperparameter configuration were used to train a final model that leverages data from six datasets. For this purpose, 25% of the data was reserved for validation and resampled at each epoch. This procedure enables the use of an early-stopping mechanism to prevent overfitting while ensuring the model extracts information from all available data.

5. RESULTS ANALYSIS

This section analyzes the results obtained with the final MLP model trained in the previous section. Figure 3 shows the predictions obtained for the training datasets compared to the actual degradation values, while Table 5 and Table 6 summarize the corresponding quantitative performance metrics, aggregated by training set and by cut interval, respectively.

Table 5. Results by training set.

Training set	Mean RMSE	Mean MAPE	Mean R^2
1	10.607	43.728	0.452
2	3.921	11.180	0.798
3	3.634	10.089	0.877
4	13.794	108.635	-0.327
5	10.018	62.097	0.206
6	5.908	35.748	0.511

Table 6. Results by cut interval.

Cut interval	Mean RMSE	Mean MAPE	Mean R^2
2-6	2.822	10.756	0.882
7-11	4.477	12.980	0.691
12-16	6.827	37.275	0.408
17-21	17.093	123.693	-0.054
22-26	4.698	41.526	0.639

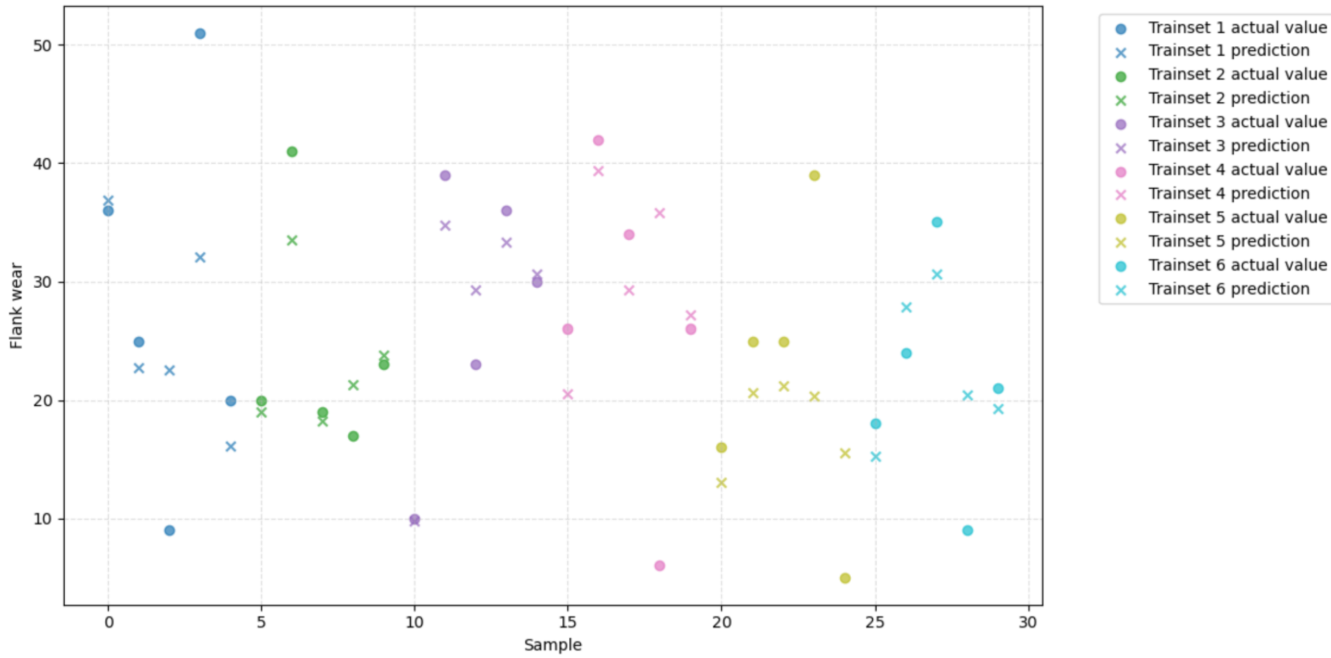
Table 2. Selected features.

Feature	Source	Sensor/Variable	Formula
DCOUNT	Controller	Z load	$\sum_i \left(\lvert x_{i+1} - x_i \rvert > 0.1(x_{max} - x_{min}) \right)$
SMED	Frequency-domain	Acceleration Y	$\text{median}\left(\frac{\lvert \text{FFT}(x) \rvert^2}{N}\right)$
XMAX	Controller	X load	$\max(x)$
XP2P	Controller	Z load	$x_{max} - x_{min}$
DMEAN	Controller	Z load	$\frac{1}{N-1}\sum_{i=1}^{N-1}(x_{i+1} - x_i)$
XCLEAR	Controller	X load	$\frac{x_{max}}{\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i \rvert}}$
XCREST	Temporal-domain	Acceleration X	$\frac{x_{max}}{\sqrt{\frac{1}{N}\sum x_i^2}}$
XKURT	Controller	Z load	$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma}\right)^4 - 3$
XCREST	Controller	X load	$\frac{x_{max}}{\sqrt{\frac{1}{N}\sum x_i^2}}$
XKURT	Temporal-domain	Acoustic emission	$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma}\right)^4 - 3$
DMEAN	Controller	Main Spindle Speed	$\frac{1}{N-1}\sum_{i=1}^{N-1}(x_{i+1} - x_i)$
XRMS	Temporal-domain	Acceleration Y	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
XSTD	Temporal-domain	Acceleration Y	$\sqrt{\frac{1}{N-1}\sum (x_i - \bar{x})^2}$
SMED	Frequency-domain	Acceleration X	$\text{median}\left(\frac{\lvert \text{FFT}(x) \rvert^2}{N}\right)$

Table 3. MLP optimization.

No. Layers	Hidden Layers Size	Activation Function	Learning Rate	Weight Decay	Batch Size	Mean RMSE	Mean MAPE	Mean R^2
3	(205, 108, 228)	ReLU	8.357×10^{-3}	1×10^{-6}	8	6.027	28.063	0.612
3	(225, 23, 201)	ReLU	9.879×10^{-3}	1×10^{-6}	8	6.068	28.096	0.608
3	(209, 137, 238)	ReLU	6.586×10^{-3}	1×10^{-6}	8	5.964	28.934	0.597
3	(218, 108, 224)	ReLU	8.810×10^{-3}	2×10^{-6}	8	6.065	29.058	0.606
3	(230, 107, 242)	ReLU	8.755×10^{-3}	2×10^{-6}	8	6.119	28.692	0.597

Figure 3. Training set results.



As shown in these results, the model exhibits substantial variations across datasets. In datasets 2 and 3, the predictions closely follow the trend of the actual values. However, this behavior is not consistent across the remaining datasets, where the model shows reduced stability. In datasets 1, 4, 5, and 6, a pronounced error is observed in the fourth prediction (corresponding to the 17–21 cut interval). This deviation is especially pronounced in dataset 4, whose feature vector was previously identified as an outlier, which explains the magnitude of the error. Overall, these results highlight heterogeneity across datasets and suggest that the inherent variability of the machining process may limit the consistency of the model.

The results obtained with this model during the evaluation process were an RMSE of 11.486, a MAPE of 8.518, and an R^2 of 0.875, placing it fourth in the competition. These results show a clearly better MAPE and R^2 compared with the training stage, but a worse RMSE. This difference can be explained by the fact that RMSE is highly sensitive to large absolute errors, whereas MAPE captures relative error and R^2 measures explained variance. Therefore, if the model produces a small number of large-magnitude errors in the evaluation set, the RMSE will be strongly penalized while the other metrics may remain stable, particularly if the remaining predictions are accurate.

6. CONCLUSIONS

This work presents a methodology designed for the PHM-AP 2025 Data Challenge. It is focused on predicting tool wear in a CNC mill-turn machine using high-frequency data with a reduced number of labeled samples. The manual feature extraction in both the temporal and frequency domains from sensor signals and controller variables, together with the feature selection process and the training of an MLP architecture, has demonstrated a competitive performance, achieving fourth place in the challenge. However, results also reveal a substantial variability across datasets and errors located in specific intervals, suggesting the need for more robust adaptive domain mechanisms. If the evaluation set is released, future work will involve evaluating other classic ML models. Furthermore, it will explore hybrid approaches, such as Deep State Space models, to improve generalization capability.

ACKNOWLEDGEMENT

Paula Mielgo's work has been funded by UVa 2023 predoctoral contracts, co-financed by Banco Santander. The work of Paula Mielgo, Anibal Bregon, and Carlos J. Alonso-González has been funded by the Spanish Ministerio de Ciencia e Innovación under grant PID2021-126659OB-I00. The authors used Grammarly to assist in improving the English language and readability of this manuscript. These tools were used solely for language editing; no AI tools were used to generate, analyze, or interpret the scientific content. were used solely for language editing; no AI tools were used to generate, analyze, or interpret the scientific content.

REFERENCES

Colantonio, L., Equeter, L., Dehombreux, P., & Ducobu, F. (2021). *A systematic literature review of cutting tool* *wear monitoring in turning by using artificial intelligence techniques*. Machines, 9(12), 351.

Hirsch, E., & Friedrich, C. (2024). *Data-driven tool wear prediction in milling, based on a process-integrated single-sensor approach.* arXiv:2412.19950 preprint.

Jones, T., & Cao, Y. (2025). *Tool wear prediction based on multisensor data fusion and machine learning*. The International Journal of Advanced Manufacturing Technology, 1–13.

Koren, Y., Ko, T.-R., Ulsoy, A. G., & Danai, K. (1991, 06). *Flank wear estimation under varying cutting conditions.* Journal of Dynamic Systems, Measurement, and Control, 113(2), 300-307. doi: 10.1115/1.2896379

Li, X. (2021). *2010 PHM Society conference data challenge.* IEEE Dataport. doi: 10.21227/jdxd-yy51

Martínez-Arellano, G., Terrazas, G., & Ratchev, S. (2019). *Tool wear classification using time series imaging and deep learning.* The International Journal of Advanced Manufacturing Technology, 104(9), 3647–3662.

Si, S., Mu, D., & Si, Z. (2024). *Intelligent tool wear prediction based on deep learning PSD-CVT model*. Scientific Reports, 14(1), 20754.

Traini, E., Bruno, G., & Lombardi, F. (2021). *Tool condition monitoring framework for predictive maintenance: a case study on milling process.* International Journal of Production Research, 59(23), 7179–7193.

Usui, E., Shirakashi, T., & Kitagawa, T. (1984). *Analytical prediction of cutting tool wear*. Wear, 100(1-3), 129–151.

Wu, D., Jennings, C., Terpenny, J., Gao, R. X., & Kumara, S. (2017). *A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests*. Journal of Manufacturing Science and Engineering, 139(7), 071018.

Xu, X., Wang, J., Zhong, B., Ming, W., & Chen, M. (2021). *Deep learning-based tool wear prediction and its application for machining process using multi-scale feature fusion and channel attention mechanism.* Measurement, 177, 109254.

Yang, Q., Pattipati, K. R., Awasthi, U., & Bollas, G. M. (2022). *Hybrid data-driven and model-informed online tool wear detection in milling machines.* Journal of manufacturing systems, 63, 329–343.

Zhang, C., Yao, X., Zhang, J., & Jin, H. (2016). *Tool condition monitoring and remaining useful life prognostic based on a wireless sensor in dry milling operations*. Sensors, 16(6), 795.

Zhang, Y., & Zhu, K. (2022). *Tool wear estimation with a data-driven physics coupling approach*. Manufacturing Letters, 34, 38–42.

Zhou, Y., Liu, C., Yu, X., Liu, B., & Quan, Y. (2022). *Tool wear mechanism, monitoring and remaining useful life (RUL) technology based on big data: A review.* SN Applied Sciences, 4(8), 232.

Zhou, Y., & Xue, W. (2018). *Review of tool condition monitoring methods in milling processes*. The International Journal of Advanced Manufacturing Technology, 96(5), 2509–2523.

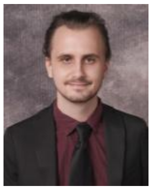
BIOGRAPHIES

Paula Mielgo is a Ph.D. Student in the Computer Science Department at the University of Valladolid, Spain. Mielgo completed her two Bachelor's Degrees in Mathematics and Computer Science Engineering in 2022. Furthermore, she obtained a Master's Degree in Computer Science Engineering in 2023. During her studies, she received several academic awards, including the award for the best academic results in both of her CS programs and a regional award for her final degree thesis. Later that year, she started her Ph.D. thesis in Computer Science.

Marcos Quinones-Grueiro received M.S. degree in industrial informatics and automation in 2017, and a Ph.D. degree in technical sciences with specialization in automation in 2018, both from the Technological University of Havana. He is currently a research scientist at the Institute for Software Integrated Systems at Vanderbilt University. Marcos' career and research interests focus on developing and applying foundational machine learning and reinforcement learning methods for the safety, control, and optimization of cyber-physical systems in crucial areas, such as transportation, water industry, manufacturing, and energy. He has authored a book, several journal papers, and book chapters, and participated in more than 35 international conferences including ACC, CDC, ICRA, CCTA, and IFAC World Congress.

Anibal Bregon received his B.Sc., M.Sc. and Ph.D. degrees in Computer Science from the University of Valladolid (Spain) in 2005, 2007 and 2010, respectively. He joined the Department of Computer Science at the University of Valladolid in 2011, where he is Associate Professor since February 2018. He has carried out both basic and applied research in the areas of fault diagnosis and prognosis for aerospace and industrial systems, and has co-authored more than 85 journal and conference papers. He is currently leading a national funded project on advanced learning for smart manufacturing and several technology transfer contracts on Deep Learning, and has also participated as researcher on several funded projects, networks and contracts on fault diagnosis and prognosis topics, on Big Data analytics and on Deep Learning. He has been guest researcher with the Intelligent Systems Division at NASA Ames Research Center and the Institute for

Software Integrated Systems at Vanderbilt University, among others. His current research interests include model-based reasoning for diagnosis and prognosis, health-management, Big Data, Industry 4.0 and Deep Learning. Among various other professional activities, he has held different chair positions at the PHM and PHME conferences, has been co-administrator of several courses and summer schools on diagnosis, prognosis, and artificial intelligence, and has been the Local Chair of the 2016 European Conference of the Prognostics and Health Management Society.



Austin Coursey is a computer science Ph.D. student in the Modeling and Analysis of Complex Systems lab at the Institute for Software Integrated Systems at Vanderbilt University. His research interests include reinforcement learning, continual learning, prognostics, and anomaly detection. He is currently researching the safety of continual reinforcement learning as part of his NSF graduate fellowship.

Carlos J. Alonso-González received the B.S. and Ph.D. degrees in physics from the University of Valladolid, in 1985 and 1990, respectively. After a brief stay in private companies and the Public University of Navarra, he joined the University of Valladolid, where he is currently an Associate Professor with the Department of Computer Science. He is also the Head of the Intelligent Systems Group, Department of Computer Science. He has worked on different national and European-funded projects related to the monitoring and diagnosis of continuous industrial environments and dynamic hybrid systems. He is also involved in projects related to the application of deep learning and casual and explainable AI to Industry 4.0, both for manufacturing and continuous processes. His current research interests include knowledge, model, and data-based systems for health management of dynamic systems, model and data-based diagnosis and prognosis of complex physical systems, and machine learning. He has been a member of the Board of Trustees of the Sugar Technology Center, being responsible for projects related to the application of artificial intelligence to online production supervision.

Gautam Biswas received B.S. degree in electrical engineering from the Indian Institute of Technoligy in Mumbai, India, and M.S. and Ph.D. degrees in computer science from Michigan State University in E.Lansing, MI. He is a Cornelius Vanderbilt Professor of Engineering, and Professor of Computer Science and Computer Engineering in the CS Department. He also is a senior research scientist at the Institute for Software Integrated Systems. Prof. Biswas conducts research in Intelligent Systems with primary interests in monitoring, control, and fault adaptivity of complex cyber physical systems. In particular, his research focuses on Deep Reinforcement Learning, Unsupervised and Semi-supervised Anomaly Detection methods, and Online Risk and Safety analysis applied to Air and Marine vehicles as well as Smart Buildings. His work, in conjunction with Honeywell Technical Center and NASA Ames, led to the NASA 2011 Aeronautics Research Mission Directorate Technology and Innovation Group Award for Vehicle Level Reasoning System and Data Mining methods to improve aircraft diagnostic and prognostic systems.