

Article

AI-Driven Tool Wear Prediction Under Severe Data Scarcity with SHAP-Guided Feature Selection and Fold-Safe Augmentation: A Case Study of Titanium Microdrilling

Saman Fattahi ^{1,*} , Bahman Azarhoushang ^{1,2,*} , Masih Paknejad ¹  and Heike Kitzig-Frank ¹

¹ KSF Institute for Advanced Manufacturing, Furtwangen University, 78532 Tuttlingen, Germany; masih.paknejad@hs-furtwangen.de (M.P.); heike.kitzig-frank@hs-furtwangen.de (H.K.-F.)

² Department of Microsystems Engineering (IMTEK), University of Freiburg, 79098 Freiburg, Germany

* Correspondence: saman.fattahi@hs-furtwangen.de (S.F.); bahman.azarhoushang@hs-furtwangen.de (B.A.)

Abstract

Microdrilling of titanium alloys suffers from rapid tool wear that degrades surface quality and dimensional accuracy, while industrial datasets are often too small for conventional data-hungry models. This work proposes a general, AI-driven modelling framework for tool wear prediction under severe data scarcity, which is validated using a titanium microdrilling case study. The study focuses on maximum flank-wear prediction ($VB_{\max}$) using 18 experimental observations ($VB_{\max} = 4\text{--}13\text{ }\mu\text{m}$). Three regression models—support vector regression (SVR), random forest (RF), and extreme gradient boosting (XGBoost)—were benchmarked under multiple validation protocols, with leave-one-out cross-validation (LOOCV) used as the primary assessment due to the limited sample size. To improve reliability and transparency, feature selection was performed using SHapley Additive exPlanations (SHAP), yielding a compact, interpretable feature subset dominated by thrust-force descriptors. Robustness was further evaluated using hyperparameter tuning and a conservative, leakage-controlled (“fold-safe”) augmentation strategy applied strictly within training folds. After tuning and fold-safe augmentation, XGBoost achieved the best LOOCV performance ($R^2 = 0.89$, $\text{MSE} = 0.70\text{ }\mu\text{m}^2$, $\text{MAPE} = 7.62\%$). External validation on two additional tools under identical cutting conditions using a frozen model configuration showed bounded prediction errors under geometry and coating shifts. Overall, the results indicate that combining systematic benchmarking, SHAP-guided explainable feature selection, and leakage-controlled augmentation can enable accurate and interpretable $VB_{\max}$ prediction in the investigated titanium microdrilling case study, while broader validation across additional tools and cutting conditions is required to confirm generalization.

Keywords: microdrilling; maximum flank wear ($VB_{\max}$); tool wear prediction; limited data; XGBoost; SHAP; fold-safe data augmentation



Academic Editor: Alexios Papacharalampopoulos

Received: 7 January 2026

Revised: 3 February 2026

Accepted: 5 February 2026

Published: 9 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The microdrilling process plays a vital role in precision manufacturing sectors such as biomedical devices, aerospace components, and electronics. The high degree of accuracy in hole dimensions and shapes is vital for the reliability and functionality of these sensitive products [1,2]. Among the various workpiece materials, titanium alloys are widely favored due to their exceptional strength-to-weight ratio, biocompatibility, and corrosion resistance [3]. However, these properties also make titanium extremely difficult to machine, particularly under micro-scale conditions [4]. Severe tool wear, chip adhesion, poor heat

dissipation, and rapid flank deterioration are typical challenges encountered in titanium microdrilling [5]. These effects not only compromise surface quality and dimensional accuracy but also limit productivity and increase the likelihood of tool breakage [6]. To address these challenges, tool wear monitoring and prediction have become crucial in ensuring the reliability and sustainability of microdrilling operations [7]. Traditional monitoring approaches—often manual and empirical—are insufficient for capturing the dynamic and non-linear behavior of wear progression at the microscale [8]. This necessitates a transition toward intelligent and autonomous solutions capable of processing real-time sensor data [9]. Within this context, data-driven manufacturing frameworks are gaining prominence, particularly in the era of Industry 4.0 [10]. In this context, machine learning (ML) has proven especially effective for tool wear prediction [11,12]. ML models can identify subtle patterns in complex datasets and support both classification and regression tasks. When applied to signals such as drilling forces, vibrations, or acoustic emissions, they offer significant potential for estimating tool wear [13–15].

The application of small datasets in data-driven machining processes is a critical area of research, as it directly addresses the practical challenges of limited data availability in industrial environments. In manufacturing, small datasets are the norm due to the high cost, time constraints, and operational disruptions associated with experimental data collection [16]. Such limitations hinder the development of accurate and robust predictive models, as small sample sizes often fail to capture the full variability and complexity inherent in machining processes. Recent studies have sought to overcome these constraints by employing innovative approaches such as advanced machine learning algorithms, data augmentation techniques, and virtual sample generation. For example, Lv et al. [17] used strategies like Cut-Flip and Mix-Normal to augment small datasets in machinery fault diagnosis, resulting in classification accuracy improvements of 10–30%. Similarly, Liu & Tian [16] used particle swarm optimization combined with broad learning systems (PSOVSGBLS), which has enabled effective virtual sample generation, reducing prediction errors in surface roughness modeling for ultra-precision machining. In related machining contexts, Siahsharani et al. [18] demonstrated that combining data-efficient learning with dimensionality reduction (e.g., PCA) and multi-criteria optimization can yield robust process insights under severe experimental data limitations.

While deep learning algorithms excel in handling complex, high-dimensional data such as images and video, they require substantial computational resources and large datasets for effective training—factors that are often prohibitive in precision manufacturing scenarios. In contrast, traditional machine learning techniques and ensemble methods, such as support vector machines and XGBoost, are not only more computationally efficient but also well-suited for small and imbalanced datasets commonly encountered in machining applications. This relationship is illustrated in Figure 1, which schematically compares the performance trends of deep learning models and traditional machine learning algorithms as a function of dataset size. As shown, deep learning algorithms generally demonstrate more progressive performance improvements with increasing data, eventually surpassing traditional ML algorithms once a critical dataset size is reached. However, for small datasets—a scenario typical in industrial manufacturing—deep learning models are often impractical due to overfitting and insufficient generalization [19]. The critical point at which deep learning begins to outperform traditional ML varies, but a common rule of thumb is that deep models require at least ten times as many data points as there are input features to avoid overfitting risk [20]. As a result, for data-limited regimes, classical machine learning and ensemble methods (e.g., support vector machines and gradient-boosted trees) are frequently preferred due to their comparatively lower sample complexity, competitive accuracy, and practical training requirements. For example, Nenchev et al. [21]

have demonstrated that XGBoost has superior performance in handling class imbalance and inhomogeneity in small datasets, outperforming traditional empirical models in tasks such as gear steel hardenability prediction. In addition, these models do not suffer from the complete ‘black-box’ nature of deep architectures and typically require less computational infrastructure for training and deployment [22]. Recent research has also highlighted the value of advanced regularization and transfer learning techniques for robust predictions with limited data. For instance, Turung et al. [23] used Bayesian regularization in neural networks and demonstrated improved stability and transferability across multiple datasets. More recently, image-based machine learning approaches have also been reported for tool wear estimation, highlighting that data-efficient and interpretable modeling strategies can be extended beyond signal-based monitoring and across different machining processes and materials [24]. These observations motivate the development of data-efficient and explainable modeling pipelines for tool wear prediction when extensive experimental datasets are not available.

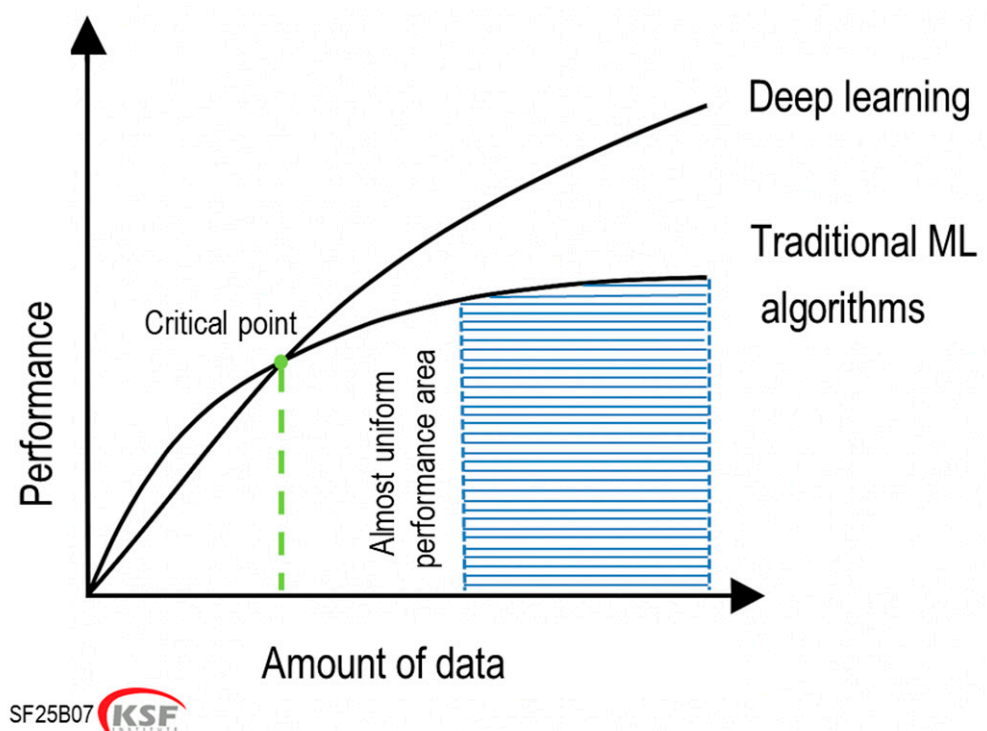


Figure 1. Influence of dataset size on model performance for deep learning and traditional machine learning algorithms. The schematic is an original illustration, conceptually inspired by [18,19].

Feature extraction and selection are critical steps in traditional ML-based tool wear prediction. The quality and relevance of extracted features directly influence predictive accuracy [8]. However, many high-performing ML models—particularly tree-based ensembles and kernel methods—offer limited transparency [25], which can hinder process validation in precision applications such as microdrilling. This has led to increasing interest in explainable AI (XAI) approaches that improve model interpretability without necessarily sacrificing accuracy [26]. Among these, SHAP (SHapley Additive exPlanations) is a widely used feature attribution method that quantifies the contribution of each input variable to a model’s prediction [27]. In tool wear prediction, SHAP primarily supports interpretability and can be used to guide feature selection; when combined with rigorous validation, SHAP-guided feature reduction can reduce feature redundancy and improve generalization in small-data settings [28,29]. The integration of interpretable feature selection with advanced ML models is therefore a promising direction for data-constrained

microdrilling wear prediction, particularly for titanium alloys under challenging cutting conditions [30]. As previously discussed, commonly used signal sources for tool wear monitoring include vibration, acoustic emission (AE), cutting forces, and vision-based data. Prior studies have demonstrated the feasibility of wear monitoring using these modalities and a range of ML models, although many investigations rely on larger datasets or focus on classification rather than regression under severe data scarcity. Cutting force signals are often favored due to their direct relationship with wear mechanisms. For instance, Varghese et al. [31] used force data to identify tool wear stages in micro-milling and demonstrated the effectiveness of ML models such as Random Forest. Vibration-based approaches have also been applied for wear monitoring, as reported by Gomes et al. [8] and Yang et al. [32]. AE signals are particularly sensitive to high-frequency events (e.g., chipping), as highlighted by Yan et al. [33]. Regarding titanium machining, Sharma et al. [34] compared several models for predicting maximum flank wear during the turning of Ti6Al4V and reported stable performance across varying cutting parameters. More broadly, studies such as Shurrab et al. [35] have shown that supervised learning can support tool-condition assessment using different sensor combinations and feature sets. Misal et al. [36] also demonstrated the potential of automated time-series feature extraction (TSFEL) for tool wear prediction and wear progression assessment, highlighting the value of combining feature engineering with ML evaluation using metrics such as MSE, MAPE, and R^2 .

While prior studies have demonstrated the effectiveness of machine learning methods for tool wear prediction, critical challenges persist, especially concerning titanium microdrilling under severely data-limited conditions. Despite extensive research, few studies have systematically compared multiple machine learning algorithms specifically regarding their capability to handle the inherent constraints of small datasets commonly encountered in precision manufacturing. Additionally, existing research has seldom rigorously evaluated various validation and data augmentation strategies explicitly designed to enhance predictive robustness under these limited data scenarios. Furthermore, interpretability and transparency in predictive modeling—crucial elements for practical industrial adoption—have received comparatively less attention, with transparent feature selection techniques such as SHAP rarely utilized within this context.

Accordingly, this study aims to (i) benchmark established regression models for predicting VB_{max} under severe data scarcity, validated here using a titanium microdrilling case study, (ii) improve interpretability and reduce feature redundancy using SHAP-guided feature selection, and (iii) assess model robustness using small-sample validation protocols with LOOCV as the primary evaluation strategy, complemented by conservative, leakage-controlled data augmentation and hyperparameter tuning. The study further examines which sensor-derived features are most influential for VB_{max} prediction in the investigated wear range, providing practical guidance for data-constrained microdrilling applications. While titanium microdrilling is used here as a challenging validation domain, the core contribution of this work is an AI-driven, data-efficient modelling framework that is not process-specific and can be transferred to other machining operations operating under similar data scarcity.

2. Models and Methodology

2.1. Machine Learning Models for Tool Wear Prediction

As previously mentioned, the primary objective of this study is to achieve quality prediction—specifically, tool wear estimation—through supervised machine learning regression with a limited dataset. Given the extensive range of available algorithms, selecting a suitable model is crucial for both predictive accuracy and practical deployability. To ensure an informed model selection, eight algorithms—covering linear models, kernel

methods, ensemble techniques, shallow neural networks, and gradient boosting—were initially screened. The evaluation followed explicit criteria widely adopted in tool-wear prediction research [37–43], including predictive accuracy, robustness under limited data, interpretability, computational efficiency, and hyperparameter complexity. Table 1 summarizes representative machine learning models reported in the literature [37–61], together with their typical applicability under different data regimes. Model selection was therefore guided by explicit criteria relevant to data-limited machining applications, which are directly reflected in the data regime, strengths/limitations, interpretability, and inclusion rationale summarized in Table 1.

Table 1. Summary of data-driven models reported in the literature for tool wear prediction, including applicability under small-data conditions and practical limitations.

Model	Typical Data Regime (Reported)	Strengths	Limitations	Interpretability	Rationale for Inclusion	Key References
Linear Regression	Medium–large datasets	Simple baseline	Inadequate for nonlinear wear behavior	High	Used as baseline only	[37,53]
Decision Tree	Medium datasets	Simple structure	High variance, unstable with small data	High	Excluded due to instability	[55,59]
Random Forest (RF)	Small–medium datasets	Robust, reduced overfitting	Limited extrapolation	Moderate	Selected as stable ensemble model	[58,60]
Support Vector Regression (SVR)	Small datasets	Effective under limited data	Kernel-dependent, tuning sensitive	Moderate	Selected for small-data robustness	[45,50]
XGBoost	Small–medium datasets	High predictive accuracy	Risk of overfitting without validation	Low–Moderate	Selected due to strong performance	[52,54]
K-Nearest Neighbors (k-NN)	Medium–large datasets	Simple implementation	Highly noise-sensitive	Low	Excluded due to noise sensitivity	[49,51]
Artificial Neural Network (ANN)	Large datasets	Flexible nonlinear modeling	Data-hungry, poor interpretability	Low	Excluded due to limited data	[47,61]
Gaussian Process Regression (GPR)	Small datasets	Uncertainty-aware	Computationally expensive	High	Excluded due to limited data	[46,48]

Models requiring large training datasets (e.g., Artificial Neural Network (ANN)), exhibiting high sensitivity to noise (e.g., K-Nearest Neighbors (k-NN)), or incurring prohibitive computational cost under increasing feature dimensionality (e.g., Gaussian Process Regression (GPR)) were excluded from further analysis. Based on these criteria, SVR, Random Forest, and XGBoost were selected for detailed benchmarking, as summarized in Table 1. SVR provides kernel-based nonlinear regression using the ϵ -insensitive loss function and radial basis function (RBF) mapping [52,56,62,63]; RF offers variance reduc-

tion and intrinsic feature-importance estimation through ensemble averaging [58,60,64,65]; and XGBoost incorporates regularized gradient boosting with strong predictive accuracy and overfitting control [52,56,66–68]. These models are well-suited for small datasets and are widely used in machining analytics. Their key hyperparameters were tuned to balance accuracy and generalization under data-limited conditions. While these models offer distinct advantages, their limitations must also be considered. SVR can be sensitive to outliers, which may be mitigated through robust loss functions. Random Forest models, although stable, may require careful tuning to avoid overfitting in high-dimensional feature spaces. XGBoost, despite its high predictive capability, demands careful hyperparameter optimization to achieve reliable generalization. Understanding these trade-offs is essential for applying these models effectively in practical, data-scarce machining scenarios.

Figure 2 presents an integrated overview of the proposed baseline benchmarking workflow used for tool-wear prediction in titanium microdrilling, adapted for data-limited conditions. For clarity, the workflow is organised into three main stages—data acquisition and screening, model development and feature engineering, and validation/tuning—while internally comprising six technical components.

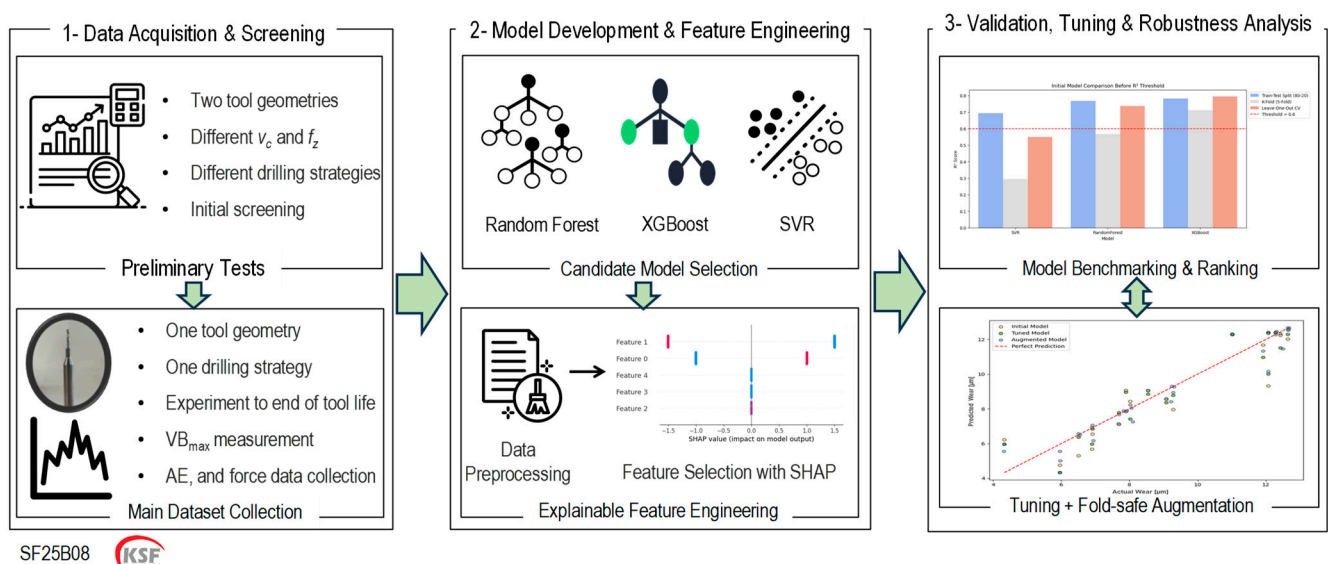


Figure 2. Overview of the proposed baseline benchmarking workflow used for tool-wear prediction in titanium microdrilling, adapted for data-limited conditions.

2.1.1. Stage 1: Data Acquisition & Screening

Stage 1 establishes the experimental foundation for the proposed framework and corresponds to the Data Acquisition & Screening block in Figure 2. Since all subsequent modeling, feature engineering, and validation steps rely on the quality and consistency of the acquired data, particular attention was given to defining stable operating conditions and a repeatable drilling strategy.

Microdrilling experiments were conducted under controlled laboratory conditions using two micro-drill geometries and multiple combinations of cutting speed and feed per tooth. An initial screening phase was performed to identify parameter ranges that enable progressive tool wear development without premature catastrophic failure. This screening step ensured that the collected data captured meaningful wear evolution rather than isolated breakage events. A peck drilling strategy [69] was adopted for all experiments, as it is widely recognized in microdrilling literature to improve chip evacuation and delay sudden tool breakage. By periodically separating cutting and non-cutting intervals, peck drilling also provides well-defined force transients, which are advantageous for signal

segmentation and subsequent tool condition analysis. This characteristic has been exploited in several thrust-force-based microdrilling studies and is particularly beneficial for data-driven wear modeling under limited data conditions.

Following the screening phase, the main dataset was collected using a single tool geometry and a fixed drilling strategy to minimize confounding effects. Drilling was continued until the end of tool life, and the maximum flank wear (VB_{max}) was periodically measured using optical microscopy. In parallel, thrust force and acoustic emission signals were continuously acquired during each drilling cycle. This protocol ensures that the resulting dataset is internally consistent, physically meaningful, and suitable for data-efficient modeling under severe sample-size constraints.

2.1.2. Stage 2: Model Development & Feature Engineering

This stage established the modelling pipeline for the baseline benchmark study, with emphasis on robustness under severe data scarcity. First, an initial screening of candidate algorithms was conducted (Section 2.1) to identify regression models that balance predictive accuracy with small-data robustness, manageable hyperparameter complexity, and practical interpretability. Based on this screening, three benchmark candidates were selected for detailed evaluation: RF, XGBoost, and SVR.

The workflow then implemented an explainable feature-engineering procedure designed for limited-data learning. This procedure comprised two main steps: (i) signal preprocessing and segmentation, applied to clean the raw force and acoustic emission time-series, suppress noise, and isolate drilling-relevant regions for consistent feature computation; and (ii) SHAP-guided explainable feature engineering, performed to derive, rank, and refine statistically and physically meaningful descriptors from the segmented signals. The resulting compact and interpretable feature set was subsequently used for training, benchmarking, and comparison of the selected regression models, with VB_{max} as the target output.

2.1.3. Stage 3: Validation, Tuning & Robustness Analysis

This stage quantified model generalization and robustness through two interconnected evaluation steps. First, the explainable, limited-data-friendly feature set obtained in Stage 2 was used for Model benchmarking & ranking of the candidate regressors. Performance was assessed under multiple validation protocols using standard regression metrics (R^2 , MAE, and MSE). The resulting metric profiles were then used to systematically compare models and to rank them based on predictive accuracy and stability under the restricted sample size.

Second, the effect of hyperparameter tuning and fold-safe, variance-controlled data augmentation was evaluated for the selected model. Augmentation was applied within training folds only to increase the diversity of the training data without introducing artificial trends or information leakage. The models were subsequently re-trained and re-evaluated to compare the initial, tuned, and augmented variants, thereby quantifying the net contribution of tuning and augmentation to generalization performance.

3. Experimentation, Data Acquisition and Screening

In this section, the Stage-1 experimental procedure for data acquisition and screening is described, including the machining setup, tooling, instrumentation, and test strategy used for tool-wear monitoring and sensor-based signal collection.

3.1. Experimental Setup

Figure 3 presents the experimental setup used to investigate tool wear during micro-drilling of titanium alloys. The experiments were performed on a high-precision 5-axis

CNC milling center (KERN Microtechnik GmbH, Eschenlohe, Germany, Figure 3a), configured with high-frequency spindles offering speeds up to 80,000 rpm (HSK 25, air bearing) and 42,000 rpm (HSK 40, roller bearing), enabling exceptional precision and flexibility for micro-machining applications requiring high dimensional accuracy and surface quality. The microdrilling tools employed were solid carbide microdrills (TD.MI.080.3D, HB microtec GmbH, Tuttlingen, Germany), selected based on manufacturer recommendations to support process stability and industrial relevance. A standard external oil-based cooling was applied during drilling. The oil was supplied externally at constant conditions and was not treated as a control variable. To monitor the process, a Kistler dynamometer (Kistler Group, Winterthur, Switzerland; Figure 3b) was integrated for in-process measurement of drilling forces, while an acoustic emission (AE) sensor (DITTEL Messtechnik GmbH, Landsberg am Lech, Germany) mounted on the workpiece (Figure 3c) enabled synchronized acquisition of AE signals during drilling. This instrumentation was essential for generating high-quality, reproducible sensor data for subsequent feature extraction and machine learning analysis. Figure 3d shows the TD.MI.080.3D microdrill used in this study (tool geometry and representative drilling parameters, which are described in Table 2).

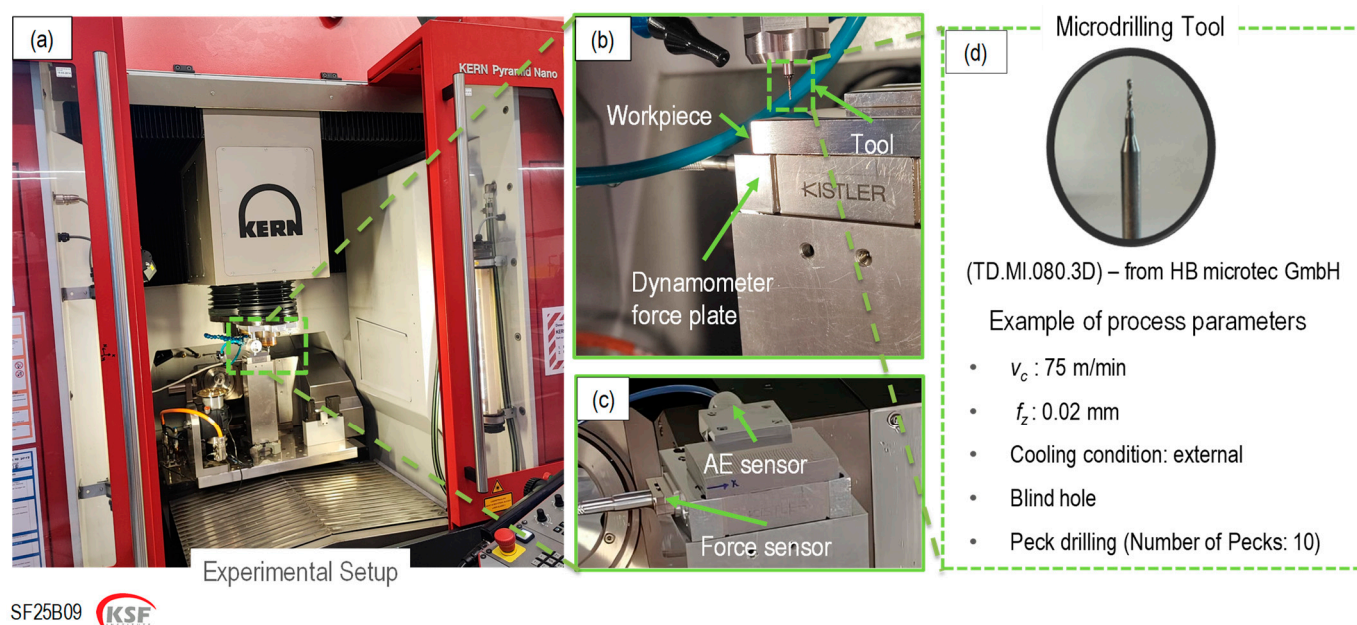


Figure 3. Experimental setup with example of process parameters (a) KERN Pyramid Nano machine, (b) workpiece, tool, and force plate setup, (c) AE and force sensor mounting on workpiece, and (d) microdrilling tool and example of process parameters.

Table 2. Machining Condition for Microdrilling.

Parameter	Condition/Value
Workpiece material	Titanium Grade 5 (Ti-6Al-4V)
Microdrill geometries	TD.MI.080.3D (HB microtec GmbH)
Tool diameter	0.8 mm
Main cutting parameters	75 m/min; f_z : 0.02 mm; Drilling depth: 2 mm (blind hole)
Cooling method	External, Oil-based
Drilling strategy	Peck drilling, 10 pecks/hole, peck depth: 0.2 mm, retraction: 0.18 mm
Measurement intervals	Tool wear measured every 20 drilled holes, equal to one cycle (18 cycles)
Data acquisition signals	Force (F_x , F_y , F_z), Acoustic Emission (AE)

3.2. Test Strategies and Data Acquisition

As we mentioned in Figure 2, prior to the main experiments, a series of preliminary tests was performed to identify stable cutting conditions and to minimize sources of uncertainty, such as stochastic built-up edge (BUE) formation and erratic wear progression. One representative example of the selected stable condition is summarized in Table 2. In this configuration, the experiments were conducted at a v_c of 75 m/min and a f_z of 0.02 mm/tooth using external cooling and blind-hole drilling. The drilling strategy employed was peck drilling, consisting of 10 pecks per hole, with each peck having a depth of 0.2 mm and a retraction length of 0.18 mm. This intermittent feed–retraction strategy supports chip evacuation and reduces local heat accumulation, thereby mitigating tool loading and the likelihood of severe BUE formation while improving hole quality.

As noted previously, although exploratory experiments were performed under multiple cutting conditions, the present manuscript intentionally restricts model development to a single selected stable condition, using one tool geometry and one coating state. This controlled design minimizes epistemic uncertainty [70] and isolates methodological effects (validation strategy, feature selection, and augmentation behavior) under extreme data scarcity, without confounding variability from changing process regimes. Additional tools are considered solely for external validation of the frozen model and are not used in any stage of model development, feature selection, or hyperparameter tuning. Each drilling experiment was conducted once per tool until tool failure. No repeated trials were performed, as the objective of this baseline study was to capture continuous wear evolution under controlled conditions rather than statistical repeatability across multiple tools. The drilling experiment was executed continuously up to tool breakage. Tool wear was evaluated in discrete drilling cycles, where each cycle corresponds to a block of 20 drilled holes. This cycle length was selected after preliminary monitoring of wear evolution, because very small $VB_{\max}$ increments at the microscale can be difficult to measure reliably over shorter intervals, and longer blocks provide a more distinguishable wear change while reducing measurement uncertainty. The tool failed during cycle 19; therefore, 18 complete cycles were retained for analysis. Accordingly, the final dataset consists of 18 observations, each comprising (i) one measured wear value ($VB_{\max}$) after a 20-hole block and (ii) one corresponding feature vector extracted from the force/AE signal batch recorded for that cycle. The measured maximum flank wear values spanned a narrow microscale range of approximately 4–13 μm across the 18 drilling cycles retained for analysis.

3.3. Flank-Wear Measurement and Screening (VB and BUE)

According to ISO 3685 [71], flank wear (VB) is defined as the local width of the wear land formed on the clearance face of a cutting tool, measured perpendicular to the cutting edge. $VB_{\max}$ denotes the maximum flank wear land width observed along the cutting edge and is commonly used as a conservative indicator of tool degradation and tool-life progression in machining studies [72]. Flank wear was quantified using a digital microscope (Keyence VHX-5000, KEYENCE DEUTSCHLAND GmbH, Frankfurt am Main, Germany, magnification range $100\times$ – $2500\times$). All flank-wear measurements were performed using optical microscopy with manual operator-based readings, and the definitions of VB and $VB_{\max}$ strictly follow ISO 3685. For each drilling cycle, local flank wear (VB) was measured at six locations along the two cutting edges; the maximum measured value was retained as $VB_{\max}$ for subsequent modeling. Figure 4 provides microscopic evidence of progressive flank wear and BUE formation during titanium microdrilling. Panels (a) and (b) show the tool's initial flank face at magnifications of $200\times$ and $1000\times$, respectively, illustrating the surface condition before drilling. Panels (c) and (d) highlight the rake face and show representative BUE formation near the cutting edge. Such built-up edges are promoted by

adhesion and localized thermal–mechanical loading, which are pronounced in titanium alloys due to their low thermal conductivity and high chemical affinity. As reported in related studies [73], BUE formation is inherently stochastic and can vary intermittently with local cutting conditions. Panel (e) illustrates the VB measurement definition used in this work.

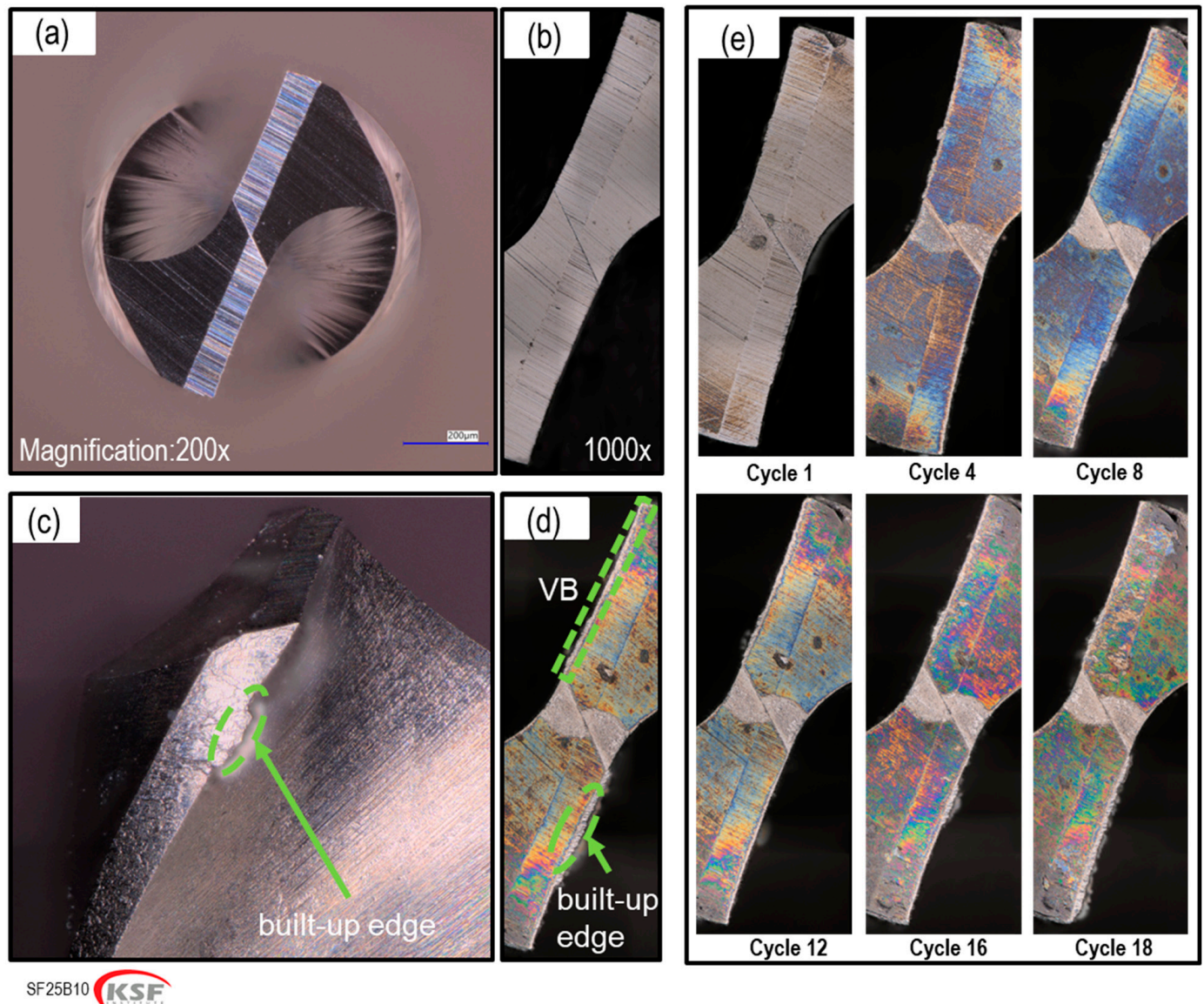


Figure 4. Microscopic images showing progressive flank wear (VB) and built-up edge (BUE) formation during microdrilling (a) Initial tool flank face at 200× magnification; (b) Initial flank face at 1000× magnification; (c) Rake face showing representative built-up edge (BUE) formation near the cutting edge; (d) Definition of flank wear width (VB) measured as a linear distance on the clearance face, with the maximum measured value along the cutting edge denoted as VB_{max} ; (e) Progressive evolution of flank wear (VB_{max}) across representative drilling cycles (cycles 1–18).

Consistent with standard machining practice, VB_{max} was used as the primary wear criterion due to its direct association with cutting performance and quality degradation. Panel (e) illustrates the visual evolution of flank wear across representative drilling cycles; with the adopted protocol of 20 holes per cycle, the shown examples correspond to cycle 1 (after 20 holes), cycle 4 (after 80 holes), cycle 8 (after 160 holes), cycle 12 (after 240 holes), cycle 16 (after 320 holes), and cycle 18 (after 360 holes). Overall, VB_{max} increases with drilling progression, indicating steady tool degradation under the selected condition.

Notably, intermittent BUE formation and detachment are visible between cycles 8 and 12: a pronounced built-up region is present at cycle 8 but is reduced/absent by cycle 12. This intermittency is consistent with observations in micromachining studies [74–76], and can induce abrupt changes in effective cutting geometry and wear rate, which complicates small-sample predictive modeling. The progressive increase in $VB_{\max}$ across cycles 1–18 therefore motivates the use of robust and interpretable feature engineering and validation procedures in the subsequent modeling sections.

4. Model Development and Validation, Results and Discussion

4.1. Signal Preprocessing and Dataset Preparation

As shown in Figure 2, this study is designed as a baseline benchmarking stage for model selection and validation under severe data scarcity. During experimentation, multiple sensor signals (F_x , F_y , F_z , and AE) were recorded to enable an initial assessment of signal quality and wear sensitivity, and to support future extensions toward multi-sensor fusion. However, for the present baseline study, signal preprocessing and modeling were intentionally restricted to the most robust and repeatable force component to ensure methodological clarity and statistical stability. Figure 5 summarises the signal preprocessing and dataset preparation pipeline applied before model development. These steps ensure consistent, machining-relevant force/AE segments for reliable feature computation and subsequent learning:

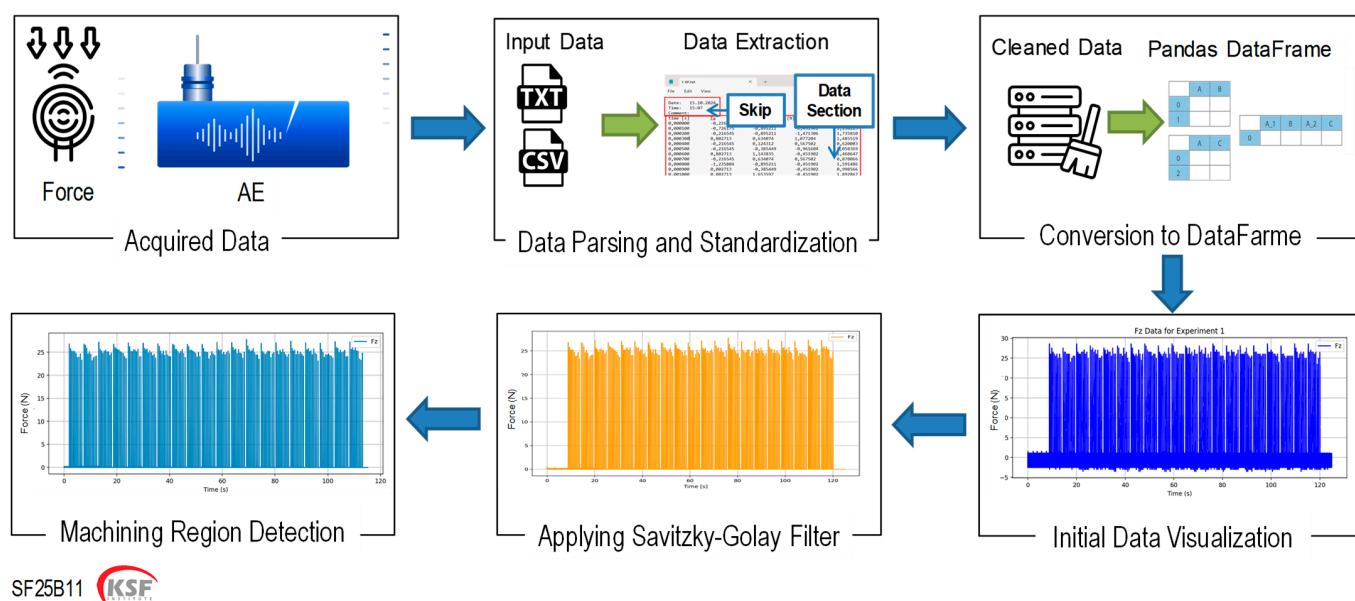


Figure 5. Data preparation for signal data.

Data parsing and standardization: Raw force data (TXT) and acoustic emission (AE) data (CSV) were first cleaned by removing corrupted/incomplete records (e.g., missing samples, non-numeric entries) and non-signal content (headers/metadata). The files were then parsed and standardized by retaining only the required columns and converting them into a consistent format and unit convention, enabling uniform downstream processing across all cycles.

Conversion to DataFrame: The standardized force and AE signals were imported into Pandas DataFrames to enable consistent indexing, channel handling, and batch processing across experimental runs.

Initial data visualization: Signals were plotted for each cycle to verify data integrity and identify anomalies such as spikes, dropouts, baseline drift, or unexpected saturation prior to filtering and segmentation.

Signal smoothing: Figure 6a shows the raw signal (as an example here, F_z) during drilling, where high-frequency noise and baseline fluctuations are visible. To suppress measurement noise while preserving drilling-related transients, a Savitzky–Golay filter was applied to each signal channel (excluding the time vector) using a window length of $w = 101$ samples and a polynomial order of $p = 3$ [77,78], yielding the smoothed signal shown in Figure 6b.

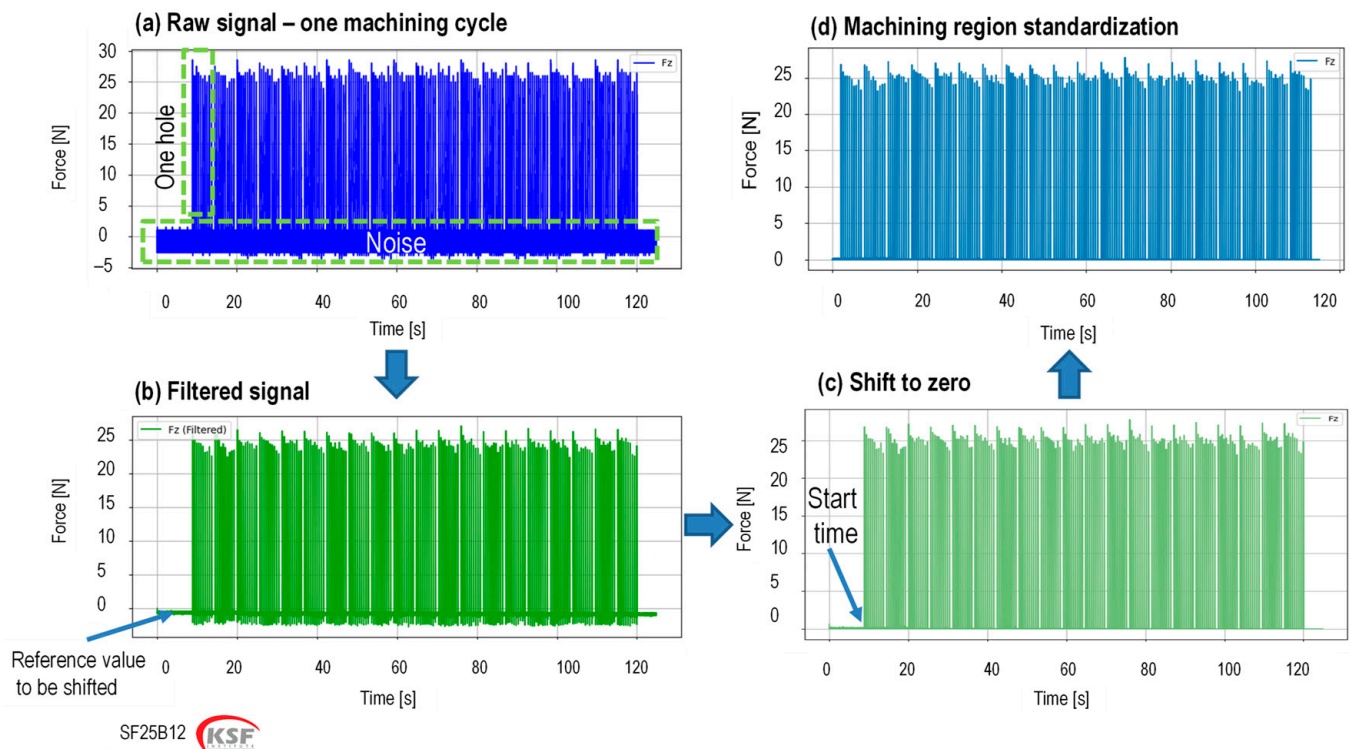


Figure 6. Multi-step force signal processing workflow: noise reduction and machining region detection.

Machining region detection: Because data acquisition was started manually, each record contains both non-cutting and cutting intervals, and the non-cutting segments vary between cycles. Figure 6c shows the smoothed F_z signal prior to segmentation. A trigger-based routine was used to automatically detect the drilling-active region (based on F_z exceeding a predefined threshold for a minimum number of consecutive samples), after which only the machining segment was retained for feature extraction (Figure 6d). This segmentation ensures that features are computed consistently from drilling-relevant intervals only.

4.2. Explainable Feature Engineering

Preliminary inspection of the acquired signals showed that the thrust force F_z exhibits the most stable behavior across drilling cycles, with lower baseline drift and a higher signal-to-noise ratio compared to the radial force components F_x and F_y . Under the selected stable drilling conditions and negligible runout, the radial forces showed near-symmetric behavior and limited sensitivity to progressive tool wear, while AE signals exhibited higher stochastic variability at the microscale. Consistent with prior microdrilling studies, thrust-force-based features—particularly F_z —have been reported to show a stronger and more monotonic relationship with flank wear progression than AE-derived indicators under comparable

conditions [69,79,80]. Therefore, because this work is positioned as a baseline/benchmark study under severe data scarcity, we intentionally restricted the modelling to F_z only to maintain a homogeneous input space and reduce the variance introduced by additional channels. Although F_x , F_y , F_z , and AE were recorded for completeness and future multi-sensor fusion studies, the baseline modeling presented in this paper relies exclusively on F_z for segmentation, feature extraction, and regression.

Feature construction and reduction were performed through an explainable, data-efficient workflow designed specifically for small-sample predictive modelling. From each segmented F_z force signal, a structured set of statistical and spectral descriptors was extracted, covering amplitude-based features (max/mean force, standard deviation, RMS, skewness, kurtosis, coefficient of variation), robust dispersion measures (MAD, IQR), peak-related metrics (number and duration of peaks), and frequency-domain indicators (spectral energy, centroid, flatness, roll-off, dominant frequency). These feature classes were selected based on prior micro-machining studies and their known sensitivity to changes in tool condition. Before feature reduction, missing entries were imputed, and all features were z-normalised to stabilise learning in the data-limited regime.

Feature selection was performed using SHapley Additive exPlanations (SHAP), which provides a model-consistent quantification of the marginal contribution of each feature to the predicted wear [81]. Rather than presenting SHAP theory [82], the focus here is on its functional role: SHAP enables (i) a consistent ranking of feature importance, (ii) inspection of directionality (whether high values increase or decrease predicted wear), and (iii) a robustness check against spurious or noise-dominated variables—critical under data-limited conditions.

SHAP-based feature ranking was performed using a dedicated reference model that is distinct from the final benchmarking models. A Random Forest regressor was selected as the reference model due to its robustness under small-sample conditions, ability to capture nonlinear relationships, and native compatibility with TreeSHAP. The reference model was trained exclusively on the training subset within each leave-one-out cross-validation (LOOCV) fold, using the segmented F_z -derived feature set and maximum flank wear ($VB_{\max}$) as the target variable. Importantly, this reference model was used solely for feature relevance estimation and dimensionality reduction and was not employed for final prediction or performance comparison. For reproducibility, the reduction criterion was defined as the mean absolute SHAP value averaged across the training samples in each validation fold, and only features showing consistently high contributions and stable sign patterns were retained. Because this study operates in an extremely small-sample regime ($n = 18$), feature dimensionality was deliberately kept low relative to n to reduce overfitting risk and improve estimator stability [83]. All preprocessing steps, including imputation, feature scaling, SHAP-based feature ranking, and any model-selection operations, were performed exclusively within the training subset of each validation fold; no information from the held-out sample was used in any stage of model construction or evaluation.

Given the extremely small sample size ($n = 18$), feature dimensionality was treated as a bias–variance trade-off and selected empirically based on cross-validated stability. We evaluated several compact SHAP-ranked feature sets and fixed the number of retained features to seven, as this provided the most stable cross-validated performance while avoiding redundancy; below seven features, performance degraded, whereas above seven, the validation error increased, indicating added noise sensitivity. The SHAP summary plot (Figure 7) shows that a small subset—primarily maximum peak force, force standard deviation, average peak force, and spectral energy—dominates predictive behaviour, while features with low and unstable SHAP influence (e.g., roll-off, dominant frequency) were excluded.

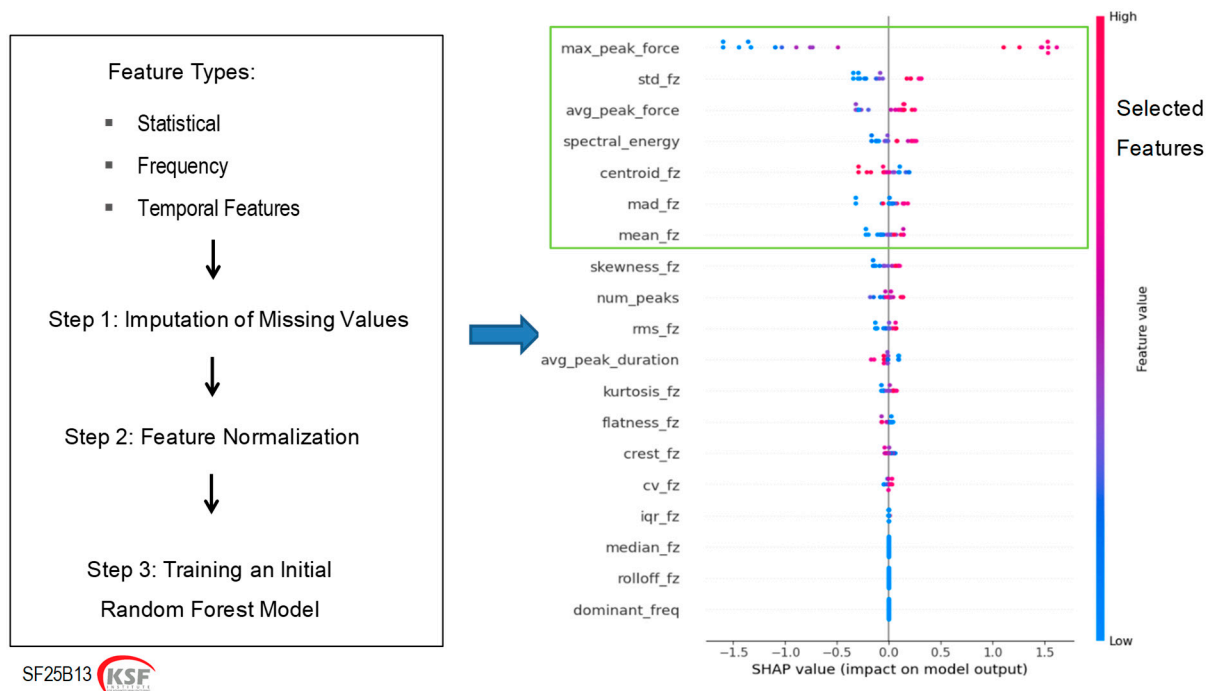


Figure 7. Feature extraction and feature importance using SHAP.

4.3. Cross-Validation Method Selection

As we already mentioned, in data-driven manufacturing, experimental datasets are often small and costly to obtain; therefore, selecting an appropriate cross-validation (CV) strategy is essential for fair and reliable assessment of model performance [84]. Given the present dataset size ($n = 18$), three complementary validation strategies were retained for benchmarking, each serving a distinct purpose. First, Leave-One-Out Cross-Validation (LOOCV) was selected as the primary protocol because it maximizes the amount of training data in each iteration ($n - 1$ samples) and ensures that every observation is tested exactly once, which is advantageous under extreme data scarcity. Second, 5-fold cross-validation was included as a variance-reducing counterpoint to LOOCV, providing a more stable estimate of generalization behavior while still reusing all samples for training and testing. Third, a conventional 80/20 train–test split was retained solely as a baseline reference aligned with common reporting practice and to facilitate comparison with prior studies, despite its known instability when the test set is very small. A broader comparison of CV variants (e.g., nested CV, stratified variants) is not reported here because the dataset is regression-based, extremely small, and the study’s focus is on the integrity of model benchmarking under data scarcity rather than on an exhaustive analysis of CV methodologies [84–93]. All candidate models were evaluated consistently under these three schemes to ensure that model ranking and performance trends were not artifacts of a single validation protocol. All regression metrics reported in this study (R^2 , MSE, MAE, and MAPE) were computed from the leave-one-out cross-validation (LOOCV) predictions. Given the dataset size ($n = 18$), LOOCV yields 18 out-of-sample predictions (one per held-out cycle), and aggregate metrics were calculated over these 18 predictions–measurement pairs.

4.4. Initial Model Selection and Screening

Model selection followed a two-step procedure. First, candidate models were screened using consistent preprocessing and the same SHAP-selected feature set. A model was excluded if it failed to achieve $R^2 \geq 0.6$ in at least two of three validation strategies (80/20 split, 5-fold CV, and LOOCV), indicating insufficient generalisation for the present

data regime. According to Figure 8, SVR did not meet this criterion, whereas Random Forest and XGBoost exceeded the threshold across all validation methods. Therefore, SVR was excluded from subsequent benchmarking, and the analysis proceeded with Random Forest and XGBoost.

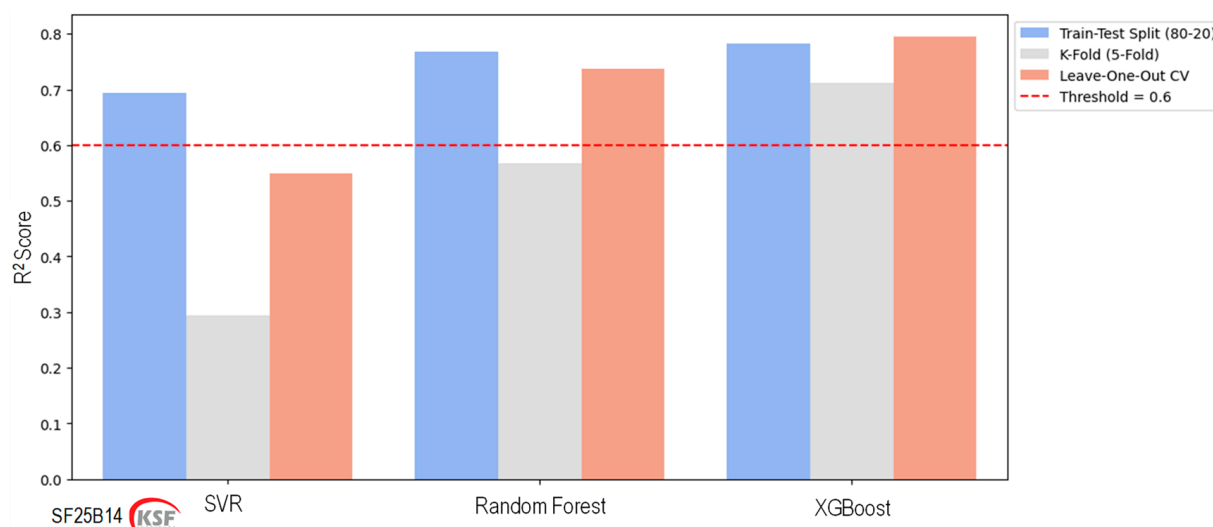


Figure 8. Comparison of SVR, Random Forest, and XGBoost performance (R^2 score) across three validation methods. The red dashed line indicates the R^2 threshold (0.6) for model selection.

4.5. Weighted Composite Score for Model Selection

To compare regression models for tool-wear prediction using multiple metrics and validation strategies, a weighted composite score was used.

Step 1: Performance metric collection

As already mentioned, for each model m and validation method, we computed four standard metrics: R^2 , MSE, MAE, and MAPE. Let $S_{m,v,p}$ denote the raw score of model m , for validation v , and performance metric p .

Step 2: metric normalization

To ensure comparability, each metric was normalized to a 0–1 range. For metrics where higher is better (e.g., R^2):

$$S'_{m,v,p} = \frac{S_{m,v,p} - S_{min}}{S_{max} - S_{min}} \quad (1)$$

For metrics where lower is better (e.g., MSE, MAE, MAPE):

$$S'_{m,v,p} = 1 - \frac{S_{m,v,p} - S_{min}}{S_{max} - S_{min}} \quad (2)$$

Step 3: weight assignment

A weight is assigned to each metric and each validation method:

w_p : weight for performance metric p (e.g., more importance to R^2 and MAPE)

w_v : weight for performance metric v (e.g., higher for LOOCV)

Step 4: Composite Score Calculation

The weighted composite score for each model is calculated as:

$$\text{Weighted composite score}_m = \sum_v w_v \left(\sum_p w_p S'_{m,v,p} \right) \quad (3)$$

where all weights are normalized so that $\sum_v w_v = 1$ and $\sum_p w_p = 1$.

Step 5: Model Ranking and Selection

Weights were assigned a priori to reflect methodological relevance under severe data scarcity. Greater emphasis was placed on LOOCV, as it maximally exploits the limited dataset and best represents the intended deployment scenario, while relative-error and variance-sensitive metrics (R^2 , MAPE, MAE) were prioritized over absolute error alone. All weights were fixed uniformly across models and validation strategies and were not optimized to avoid selection bias. Model ranking was performed using the resulting weighted composite scores, with only models exceeding the screening criterion ($R^2 > 0.6$ in at least two validation protocols) retained for final comparison. Figure 9 summarizes the weighted scores obtained under LOOCV, 5-fold cross-validation, and a conventional 80–20 train–test split. Both XGBoost and random forest models satisfy the screening requirements; however, XGBoost achieves consistently higher composite scores across all validation strategies. This result indicates superior robustness and predictive stability under the applied evaluation framework, and therefore, XGBoost is selected as the reference model for subsequent analysis in this study.

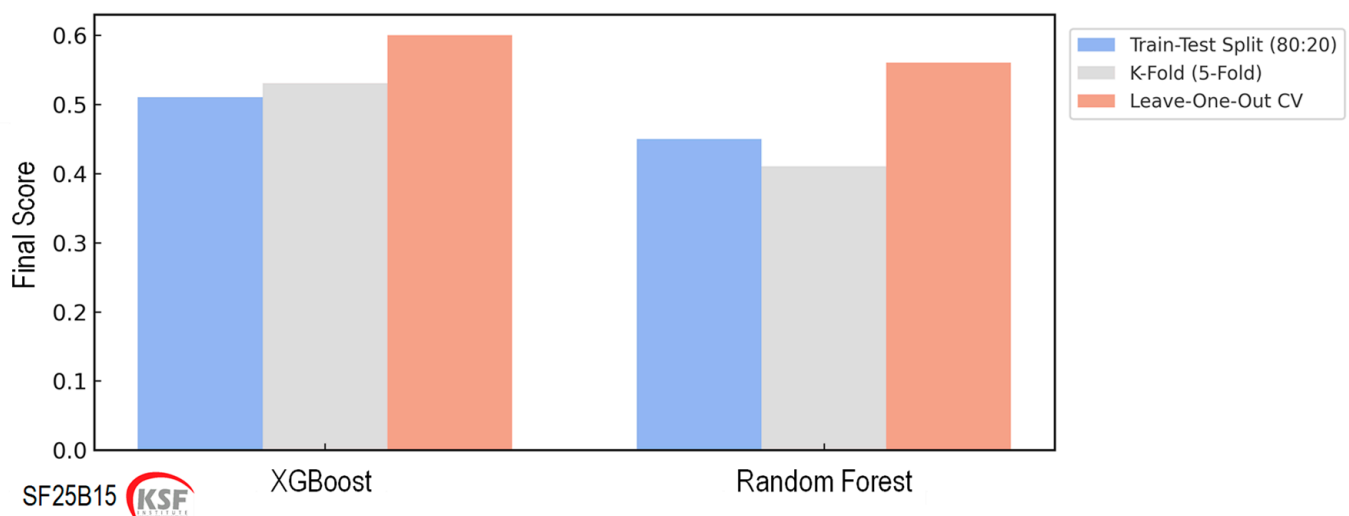


Figure 9. Model comparison based on weighted composite score.

4.6. Hyperparameter Tuning Fold-Safe Augmentation Within LOOCV (Leakage-Controlled)

Based on the weighted composite ranking (Section 4.5), XGBoost evaluated with Leave-One-Out Cross-Validation (LOOCV) was selected as the primary model–validation setting for this baseline study. Given the extremely limited sample size ($n = 18$) and the fact that each observation corresponds to one drilling cycle, summarized by time-series-derived features, robustness improvements must be introduced without compromising validation integrity. Therefore, two complementary measures were applied: (i) hyperparameter tuning and (ii) fold-safe (leakage-controlled) data augmentation embedded inside LOOCV.

4.6.1. Hyperparameter Tuning

XGBoost hyperparameters were optimized using a randomized search, where the candidate distributions for the number of estimators, learning rate, maximum tree depth, sub-sample ratio, and column sampling ratio were defined using `scipy.stats` (SciPy library [94]). `RandomizedSearchCV` from `scikit-learn` [95] with LOOCV and 50 randomly sampled configurations. This choice evaluates candidate configurations under the same extreme small-sample regime as the final benchmark and avoids the optimistic bias associated with a single hold-out split. The best-performing configuration (according to the selected scoring metric) was retained and used as the fixed (“tuned”) reference model for the subsequent augmentation study.

4.6.2. Fold-Safe (Leakage-Controlled) Data Augmentation

To mitigate estimator variance under $n = 18$ while preventing information leakage, augmentation was performed per LOOCV fold using only training samples. In LOOCV fold k , the held-out test set is $\mathcal{D}_{\text{test}}^{(k)} = \{(x_k, y_k)\}$ and the training set is $\mathcal{D}_{\text{train}}^{(k)} = \{(x_i, y_i)\}_{i \neq k}$ where x_i denotes the feature vector computed from the segmented drilling signal (here F_z), and y_i is the measured wear response $VB_{\max,i}$. Augmentation is applied only to $\mathcal{D}_{\text{train}}^{(k)}$ and never uses (x_k, y_k) or any statistics computed from the full dataset.

Within fold k , additional training samples are generated by a conservative Gaussian-bootstrap perturbation in feature space [96]. In other words, raw time-series signals are not augmented; instead, augmentation is applied to the extracted feature vectors. First, bootstrap indices π_j are sampled uniformly with replacement from the training index set $\{i: i \neq k\}$. For an augmentation multiplier m , the number of synthetic samples is

$$n_{\text{aug}} = \left\lfloor m \mid \mathcal{D}_{\text{train}}^{(k)} \right\rfloor = \lfloor m(n-1) \rfloor \quad (4)$$

In this study, the augmentation multiplier was set to $m = 1$, resulting in one synthetic sample per original training observation within each fold. Let $\sigma^{(k)} = (\sigma_1^{(k)}, \dots, \sigma_d^{(k)})$ be the feature-wise standard deviations computed only from the training fold. Each augmented feature vector is formed as the feature-wise standard deviations computed only from the training fold. Each augmented feature vector is formed as:

$$\tilde{\mathbf{x}}_j^{(k)} = \mathbf{x}_{\pi_j}^{(k)} + \boldsymbol{\varepsilon}_j^{(k)} \quad (5)$$

$$\boldsymbol{\varepsilon}_j^{(k)} \sim \mathcal{N}\left(\mathbf{0}, \alpha^2 \text{diag}\left(\left(\sigma^{(k)}\right)^2\right)\right) \quad (6)$$

where α is a small noise factor (here $\alpha = 0.02$, i.e., 2% of the training-fold feature scale). In this conservative implementation, labels are kept unchanged:

$$\tilde{y}_j^{(k)} = y_{\pi_j}^{(k)} \quad (7)$$

Label invariance under small feature perturbations is justified here because $VB_{\max}$ is measured after each full drilling cycle, while the applied perturbations ($\alpha = 2\%$) remain within the observed intra-cycle signal variability and measurement uncertainty of the extracted descriptors.

The fold-augmented training set is then

$$\mathcal{D}_{\text{train,aug}}^{(k)} = \mathcal{D}_{\text{train}}^{(k)} \cup \left\{ \left(\tilde{\mathbf{x}}_j^{(k)}, \tilde{y}_j^{(k)} \right) \right\}_{j=1}^{n_{\text{aug}}} \quad (8)$$

The tuned XGBoost model is trained on $\mathcal{D}_{\text{train,aug}}^{(k)}$ and evaluated on the single held-out sample (x_k, y_k) . This procedure is repeated for all $k = 1, \dots, n$ to obtain LOOCV predictions and aggregate metrics.

In small-sample machining datasets, a common pitfall is to apply augmentation before cross-validation or to compute augmentation scales (and other preprocessing parameters) on the full dataset, which can leak information from the held-out sample into the training distribution and yield optimistic performance estimates [97]. The fold-safe design avoids this by estimating feature-wise noise scales exclusively from $\mathcal{D}_{\text{train}}^{(k)}$ and ensuring the test observation remains unseen during augmentation. The augmentation is intentionally conservative, perturbing feature vectors only within the variability bounds estimated from the training fold, thereby increasing effective training diversity without introducing non-physical patterns. Combined with SHAP-guided feature reduction and the multi-

protocol validation strategy, the pipeline provides a reproducible benchmark framework for tool-wear regression under limited data.

4.7. Tuned, Leakage-Controlled, Augmented Selected XGBoost Model Performance

Figures 10–12 summarise the LOOCV performance of the initial, tuned, and tuned + augmentation XGBoost models. Figure 10 reports aggregate metrics (R^2 , MSE, MAE), showing a stepwise improvement from the initial baseline to the tuned configuration and further to the tuned model trained with leakage-controlled augmentation. Figure 11 compares predicted versus measured VB_{max} ; points closer to the 1:1 reference line indicate improved agreement. Figure 12 shows cycle-wise predictions across the 18 drilling cycles, where the tuned + augmentation model tracks the measured VB_{max} evolution most closely overall, noting that local non-monotonic variations reflect measurement effects associated with transient BUE formation rather than physical wear reversal.

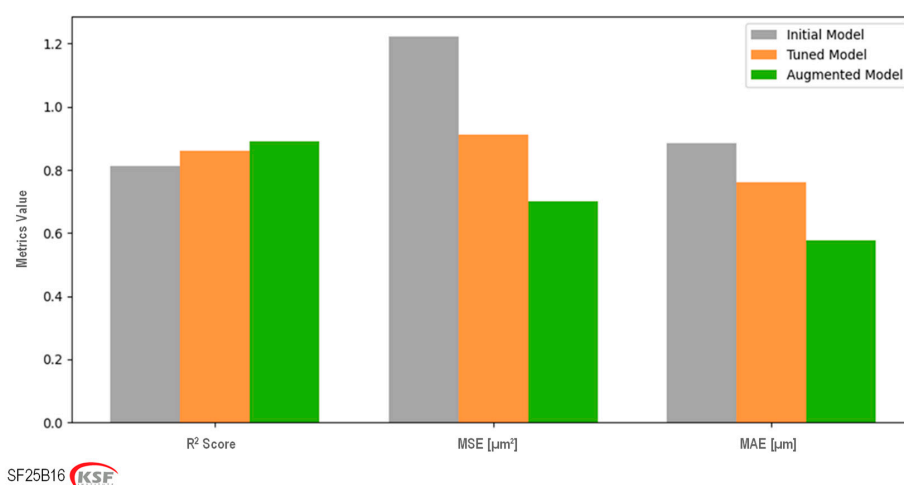


Figure 10. Comparison of LOOCV evaluation metrics (R^2 , MSE, MAE) for initial, tuned, and augmented XGBoost models.

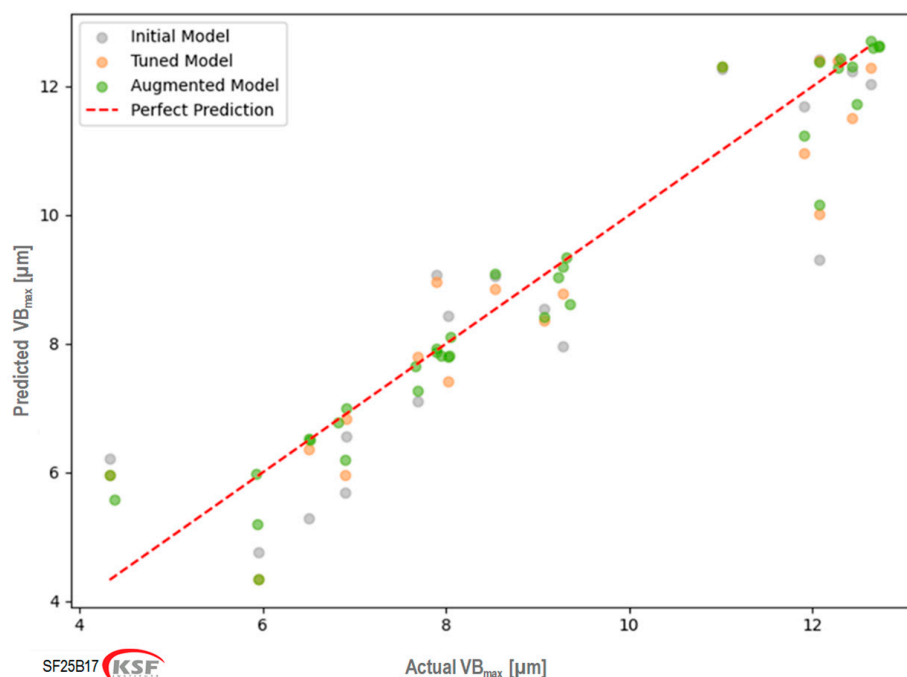


Figure 11. Actual versus predicted VB_{max} values for initial, tuned, and augmented XGBoost models using LOOCV.

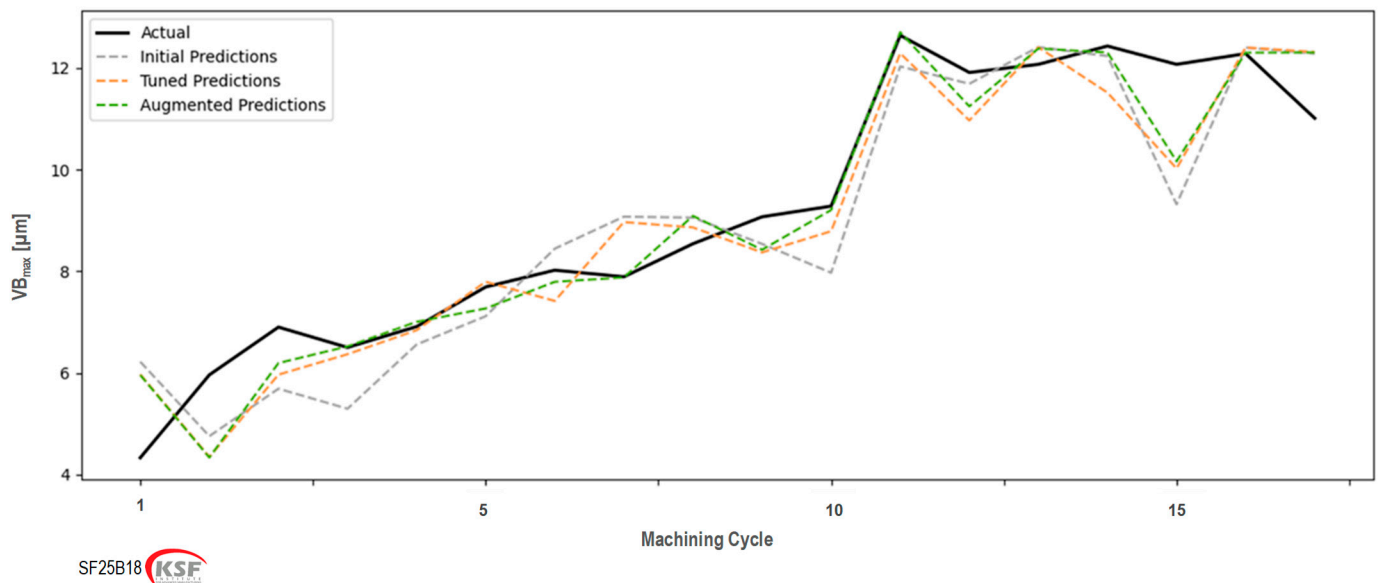


Figure 12. Machining cycle-wise comparison of measured and predicted VB_{max} values for initial, tuned, and augmented XGBoost models under LOOCV. Local non-monotonic variations are attributed to intermittent BUE formation and detachment.

Table 3 summarizes the quantitative improvements achieved at each stage. The tuned with data augmentation model achieves the best metrics, including the highest R^2 (0.892, +9.92% improvement over the initial model) and the lowest MSE (0.702, −42.63% reduction), MAE (0.578, −34.68% reduction), and MAPE (7.62%, −33.21% reduction). Notably, the data augmentation strategy delivered substantial improvements across all performance metrics, with the most significant gain being the 42.63% reduction in MSE, indicating enhanced reliability for critical predictions. These results confirm that sequential application of hyperparameter tuning and safe data augmentation provides robust, generalizable improvements in tool wear prediction for data-limited manufacturing applications, reducing prediction errors by over one-third while explaining nearly 10% more variance in tool wear behavior. While LOOCV may occasionally overestimate performance in small datasets, its use in combination with multi-validation screening, constrained hyperparameter tuning, and physically safe augmentation substantially mitigates this risk and supports the robustness of the reported results. Because the 18 observations correspond to sequential wear cycles on a single tool, strict statistical independence cannot be assumed; future work will incorporate time-aware validation schemes (e.g., forward-chaining) to explicitly address temporal dependence.

Table 3. LOOCV performance comparison of XGBoost models in different conditions.

Metric	Initial Model	Tuned Model	Tuned with Data Augmentation	Relative Improvement
R^2	0.81	0.85	0.89	9.92%
MSE	1.22	0.91	0.70	42.63%
MAE	0.88	0.76	0.57	34.68%
MAPE	11.40	9.52	7.62	33.21%

A closer inspection of Figure 12 indicates local underestimation at specific cycles (notably cycles 2 and 15). These deviations align with microscopy evidence of transient physical events (Figure 13), including built-up edge (BUE) formation and local edge dam-

age. Accordingly, local decreases in measured $VB_{\max}$ should be interpreted as transient changes in adhered material and effective edge geometry, not as a reduction in cumulative tool wear. Such events can abruptly change both the measured $VB_{\max}$ and the underlying force response, violating the implicit assumption of smooth, monotonic wear–signal evolution that many regression models rely on. Moreover, BUE formation and stochastic detachment can introduce discontinuities between consecutive cycles, producing rare patterns that are sparsely represented in an $n = 18$ training set. These observations motivate future work toward incorporating explicit physics-informed descriptors and/or uncertainty quantification to better handle low-frequency but high-impact wear mechanisms.

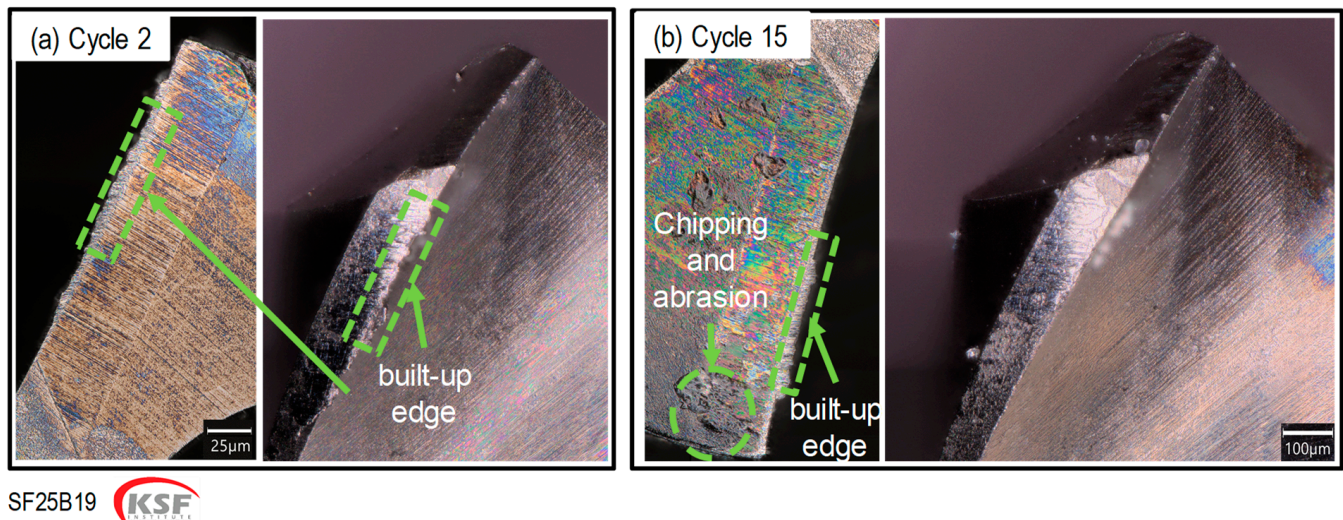


Figure 13. Microscopic images illustrating wear mechanisms linked to prediction uncertainty: (a) Cycle 2, showing prominent built-up edge on the flank face; (b) Cycle 15, highlighting both the built-up edge and local chipping/abrasion.

4.8. External Validation on Unseen Tool Geometries

To further evaluate the robustness of the proposed modeling framework beyond the single-tool development case, an external validation was conducted using two additional microdrills operated under the same cutting conditions and measurement protocol as described in Section 3.1. In this validation, the tuned XGBoost model developed in Section 4.7—comprising SHAP-based feature selection, leakage-controlled data augmentation, and optimized hyperparameters—was applied in a fully frozen configuration. All feature definitions, scaling parameters, and model settings were retained from the training case. No retraining, hyperparameter tuning, feature re-selection, or data augmentation was performed for the additional tools. Both validation tools were supplied by the same manufacturer in order to limit supplier-related variability. The first validation tool, TD.MI.080.8D, differs from the training tool primarily in its length-to-diameter ratio (8D versus 3D), introducing a geometry-induced change in tool stiffness, deflection behavior, and force dynamics. The second validation tool, TD.MI.080.3D.1, shares the same geometry as the training tool but incorporates an α -INOX coating, allowing the isolated assessment of coating effects on wear progression and tool life under otherwise identical conditions.

The resulting prediction performance is summarized in Table 4. Under the same cutting condition, the uncoated 8D tool reached a tool life of 16 cycles, whereas the α -INOX-coated 3D tool achieved 29 cycles, indicating a substantial improvement in wear resistance due to the coating. The shorter life of the 8D tool is consistent with increased compliance and higher susceptibility to deflection-induced wear mechanisms, while the extended life of the coated tool reflects reduced adhesion and improved flank wear resistance. As expected

for a strict out-of-sample evaluation, prediction errors increase relative to the internal LOOCV results obtained for the training tool. Nevertheless, the frozen model maintains bounded errors for both validation tools, with MAE values remaining below 1 μm and R^2 values exceeding 0.83. The geometry-induced shift leads to a more pronounced degradation in accuracy compared to the coating-induced shift, which preserves a closer correspondence to the training case. This behaviour indicates that while tool geometry strongly influences force signatures and wear dynamics, the learned wear–signal relationship retains functional relevance across controlled tool variations.

Table 4. External validation performance of the frozen tuned + augmented XGBoost model under identical cutting conditions.

Tool ID	Geometry/Coating	Cycles (Sample Size)	R^2	MSE [μm^2]	MAPE [%]	MAE [μm]
TD.MI.080.3D	3D, uncoated	18	0.89	0.70	7.62	0.57
TD.MI.080.8D	8D, uncoated	16	0.83	1.44	9.12	0.91
TD.MI.080.3D.1	3D, α -INOX coated	29	0.87	1.06	8.26	0.75

Overall, these results support the methodological premise of this study: when combined with leakage-aware validation, explainable feature selection, and constrained data augmentation, data-driven wear prediction models can remain predictive under severe data scarcity, even when applied to unseen tools operated under identical cutting conditions. At the same time, the observed performance differences highlight the influence of geometry and coating effects and motivate future extensions toward multi-tool training, geometry-aware descriptors, and uncertainty-informed inference.

5. Conclusions

This study presents a baseline, AI-driven, data-efficient, and interpretable framework for predicting maximum flank wear (VB_{max}) in titanium microdrilling under severe data scarcity, validated using a titanium microdrilling case study. An initial screening of eight candidate regressors led to the selection of SVR, Random Forest (RF), and XGBoost for detailed benchmarking. To improve transparency and reduce overfitting risk in the extremely small-sample regime, SHAP was used for explainable feature reduction, yielding a compact set of seven thrust-force (F_z)-derived descriptors. Using consistent preprocessing and multi-protocol validation, XGBoost showed the most stable and accurate performance and outperformed RF, while SVR did not satisfy the defined screening criterion.

For the selected XGBoost model, robustness was further improved through randomized hyperparameter tuning and a conservative, leakage-controlled (“fold-safe”) augmentation procedure embedded within LOOCV. Under this setting ($n = 18$), the tuned and fold-safe augmented XGBoost achieved the best overall performance with $R^2 = 0.89$ and reduced MSE by 42.63% relative to the initial baseline model. These results indicate that combining systematic benchmarking, SHAP-guided explainable feature selection, and fold-safe augmentation can substantially improve wear prediction accuracy even when only a very limited number of labeled drilling cycles is available. In addition, external validation on two unseen tools operated under identical cutting conditions, using a fully frozen model configuration, showed bounded prediction errors under both geometry and coating shifts, with geometry changes leading to a larger degradation than coating changes. This validation confirms bounded generalization, although broader validation across operating conditions remains future work. Although the experimental results are grounded in titanium microdrilling, the proposed workflow is not process-specific and can

be transferred to other machining operations, provided that additional external validation is conducted.

The main conclusions of this study can be summarized as follows:

- A data-efficient and interpretable AI framework enables reliable $VB_{\max}$ prediction under extreme data scarcity.
- SHAP-based feature reduction improves model stability while preserving physical interpretability.
- XGBoost outperforms SVR and RF when combined with fold-safe augmentation and controlled tuning.
- External validation confirms bounded generalization, with higher sensitivity to tool geometry than coating.
- The proposed workflow is transferable to other machining processes with appropriate validation.

6. Future Research Direction

The main limitation of the present work is the small and homogeneous dataset (single cutting condition, one tool/coating), which restricts the scope of generalization claims. Future work will therefore expand this baseline study to make a system and expanding dataset to cover a broader range of cutting conditions, tool geometries/coatings, and process parameters, enabling more rigorous assessment of robustness and transferability across realistic operating variability. In parallel, additional sensing channels (e.g., F_x , F_y , and AE) will be incorporated alongside F_z to better capture non-stationary and stochastic events such as built-up edge (BUE) formation/detachment and local edge damage, ideally within uncertainty-aware modelling frameworks that can represent confidence under rare events. Beyond extending the sensing and validation space, physics-informed learning offers a promising route to embed mechanistic constraints and domain knowledge into the predictive model, potentially improving both accuracy and physical interpretability.

Author Contributions: S.F.: Conceptualization, Software, Formal analysis, Investigation, Methodology, Data Curation, Validation, Visualization and Writing—original draft. B.A.: Conceptualization, Funding acquisition, Writing—review & editing, Supervision, Resources and Project administration. M.P.: Experimentation and Technical Consultancy. H.K.-F.: Writing—review & editing and Resources. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Ministerium für Wirtschaft, Arbeit und Tourismus Baden-Württemberg through the Invest BW program under the project “KiBoopt—Optimierung von Bohrwerkzeugen mit Methoden der KI” (Project No. BW1_4004/02).

Data Availability Statement: The data presented in this study are not publicly available due to institutional restrictions and ongoing research activities. Derived features and methodological details necessary to reproduce the analysis are fully described within the article. Data may be made available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to express their sincere gratitude for the financial support from the Ministerium für Wirtschaft, Arbeit und Tourismus Baden-Württemberg (Invest BW). They also thank HB microtec GmbH & Co. KG for providing the microdrilling tools and for their valuable technical support throughout the project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hyacinth Suganthi, X.; Natarajan, U.; Ramasubbu, N. A Review of Accuracy Enhancement in Microdrilling Operations. *Int. J. Adv. Manuf. Technol.* **2015**, *81*, 199–217. [[CrossRef](#)]
2. Chaudhary, K.; Haribhakta, V.K. Micro-Drilling on Shape Memory Alloys—A Review. *MethodsX* **2024**, *13*, 102968. [[CrossRef](#)]

3. Niinomi, M. Mechanical Properties of Biomedical Titanium Alloys. *Mater. Sci. Eng. A* **1998**, *243*, 231–236. [\[CrossRef\]](#)
4. Hourmand, M.; Sarhan, A.A.D.; Sayuti, M.; Hamdi, M. A Comprehensive Review on Machining of Titanium Alloys. *Arab. J. Sci. Eng.* **2021**, *46*, 7087–7123. [\[CrossRef\]](#)
5. Löffler, F. Wear and Cutting Performance of Coated Microdrills. *Surf. Coat. Technol.* **1998**, *107*, 191–196. [\[CrossRef\]](#)
6. Christiand, C.; Kiswanto, G.; Baskoro, A.S.; Hasymi, Z.; Ko, T.J. Tool Wear Monitoring in Micro-Milling Based on Digital Twin Technology with an Extended Kalman Filter. *J. Manuf. Mater. Process.* **2024**, *8*, 108. [\[CrossRef\]](#)
7. Beruvides, G.; Quiza, R.; Del Toro, R.; Haber, R.E. Sensing Systems and Signal Analysis to Monitor Tool Wear in Microdrilling Operations on a Sintered Tungsten–Copper Composite Material. *Sens. Actuators A Phys.* **2013**, *199*, 165–175. [\[CrossRef\]](#)
8. Gomes, M.C.; Brito, L.C.; Bacci Da Silva, M.; Viana Duarte, M.A. Tool Wear Monitoring in Micromilling Using Support Vector Machine with Vibration and Sound Sensors. *Precis. Eng.* **2021**, *67*, 137–151. [\[CrossRef\]](#)
9. Wang, S.-M.; Tsou, W.-S.; Huang, J.-W.; Chen, S.-E.; Wu, C.-C. Development of a Method and a Smart System for Tool Critical Life Real-Time Monitoring. *J. Manuf. Mater. Process.* **2024**, *8*, 194. [\[CrossRef\]](#)
10. Fattahi, S.; Azarhoushang, B.; Kitzig-Frank, H. Knowledge-Based Adaptive Design of Experiments (KADoE) for Grinding Process Optimization Using an Expert System in the Context of Industry 4.0. *J. Manuf. Mater. Process.* **2025**, *9*, 62. [\[CrossRef\]](#)
11. Wu, D.; Jennings, C.; Terpenney, J.; Gao, R.X.; Kumara, S. A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. *J. Manuf. Sci. Eng.* **2017**, *139*, 071018. [\[CrossRef\]](#)
12. Axinte, D.; Gindy, N. Assessment of the Effectiveness of a Spindle Power Signal for Tool Condition Monitoring in Machining Processes. *Int. J. Prod. Res.* **2004**, *42*, 2679–2691. [\[CrossRef\]](#)
13. Omole, S.; Dogan, H.; Lunt, A.J.G.; Kirk, S.; Shokrani, A. Using Machine Learning for Cutting Tool Condition Monitoring and Prediction during Machining of Tungsten. *Int. J. Comput. Integr. Manuf.* **2024**, *37*, 747–771. [\[CrossRef\]](#)
14. Alajmi, M.S.; Almeshal, A.M. Predicting the Tool Wear of a Drilling Process Using Novel Machine Learning XGBoost-SDA. *Materials* **2020**, *13*, 4952. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Zhou, Y.; Liu, C.; Yu, X.; Liu, B.; Quan, Y. Tool Wear Mechanism, Monitoring and Remaining Useful Life (RUL) Technology Based on Big Data: A Review. *SN Appl. Sci.* **2022**, *4*, 232. [\[CrossRef\]](#)
16. Liu, X.; Chen, G.; Li, Y.; Chen, L.; Meng, Q.; Mehdi-Souzani, C. Sampling via the Aggregation Value for Data-Driven Manufacturing. *Natl. Sci. Rev.* **2022**, *9*, nwac201. [\[CrossRef\]](#)
17. Lv, H.; Chen, J.; Zhang, T.; Hou, R.; Pan, T.; Zhou, Z. SDA: Regularization with Cut-Flip and Mix-Normal for Machinery Fault Diagnosis under Small Dataset. *ISA Trans.* **2021**, *111*, 337–349. [\[CrossRef\]](#)
18. Siahsharani, A.; Fattahi, S.; Alinaghizadeh, A.; Azarhoushang, B.; Börsinger, R. Data-Driven Optimization of Processing Parameters and Cooling Strategies in UHMWPE High Speed Milling Through Multi-Criteria Decision Making Using PCA and Pareto-Based Evolutionary Algorithms. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2026**. [\[CrossRef\]](#)
19. Ahmed, S.F.; Alam, M.S.B.; Hassan, M.; Rozbu, M.R.; Ishtiaq, T.; Rafa, N.; Mofijur, M.; Shawkat Ali, A.B.M.; Gandomi, A.H. Deep Learning Modelling Techniques: Current Progress, Applications, Advantages, and Challenges. *Artif. Intell. Rev.* **2023**, *56*, 13521–13617. [\[CrossRef\]](#)
20. Alwosheel, A.; Van Cranenburgh, S.; Chorus, C.G. Is Your Dataset Big Enough? Sample Size Requirements When Using Artificial Neural Networks for Discrete Choice Analysis. *J. Choice Model.* **2018**, *28*, 167–182. [\[CrossRef\]](#)
21. Nenchev, B.; Tao, Q.; Dong, Z.; Panwisawas, C.; Li, H.; Tao, B.; Dong, H. Evaluating Data-Driven Algorithms for Predicting Mechanical Properties with Small Datasets: A Case Study on Gear Steel Hardenability. *Int. J. Miner. Metall. Mater.* **2022**, *29*, 836–847. [\[CrossRef\]](#)
22. Domínguez-Monferrer, C.; Fernández-Pérez, J.; De Santos, R.; Miguélez, M.H.; Cantero, J.L. Machine Learning Approach in Non-Intrusive Monitoring of Tool Wear Evolution in Massive CFRP Automatic Drilling Processes in the Aircraft Industry. *J. Manuf. Syst.* **2022**, *65*, 622–639. [\[CrossRef\]](#)
23. Truong, T.T.; Airao, J.; Hojati, F.; Ilvig, C.F.; Azarhoushang, B.; Karras, P.; Aghababaei, R. Data-Driven Prediction of Tool Wear Using Bayesian Regularized Artificial Neural Networks. *Measurement* **2024**, *238*, 115303. [\[CrossRef\]](#)
24. Truong, T.T.; Airao, J.; Fattahi, S.; Azarhoushang, B.; Karras, P.; Aghababaei, R. Image-Based Machine Learning Model for Tool Wear Estimation in Milling Inconel 718. *Wear* **2025**, *571*, 205865. [\[CrossRef\]](#)
25. Gilpin, L.H.; Bau, D.; Yuan, B.Z.; Bajwa, A.; Specter, M.; Kagal, L. Explaining Explanations: An Overview of Interpretability of Machine Learning. In *Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–3 October 2018*; IEEE: New York, NJ, USA, 2018; pp. 80–89.
26. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [\[CrossRef\]](#)
27. Minh, D.; Wang, H.X.; Li, Y.F.; Nguyen, T.N. Explainable Artificial Intelligence: A Comprehensive Review. *Artif. Intell. Rev.* **2022**, *55*, 3503–3568. [\[CrossRef\]](#)
28. Alomari, Y.; Andó, M. SHAP-Based Insights for Aerospace PHM: Temporal Feature Importance, Dependencies, Robustness, and Interaction Analysis. *Results Eng.* **2024**, *21*, 101834. [\[CrossRef\]](#)

29. Cheng, W.-N.; Cheng, C.-C.; Lei, Y.-H.; Tsai, P.-C. Feature Selection for Predicting Tool Wear of Machine Tools. *Int. J. Adv. Manuf. Technol.* **2020**, *111*, 1483–1501. [\[CrossRef\]](#)
30. Shen, Y.; Yang, F.; Habibullah, M.S.; Ahmed, J.; Das, A.K.; Zhou, Y.; Ho, C.L. Predicting Tool Wear Size across Multi-Cutting Conditions Using Advanced Machine Learning Techniques. *J. Intell. Manuf.* **2021**, *32*, 1753–1766. [\[CrossRef\]](#)
31. Varghese, A.; Kulkarni, V.; Joshi, S.S. Tool Life Stage Prediction in Micro-Milling From Force Signal Analysis Using Machine Learning Methods. *J. Manuf. Sci. Eng.* **2021**, *143*, 054501. [\[CrossRef\]](#)
32. Yang, Z.; Li, L.; Zhang, Y.; Jiang, Z.; Liu, X. Tool Wear State Monitoring in Titanium Alloy Milling Based on Wavelet Packet and TTAO-CNN-BiLSTM-AM. *Processes* **2024**, *13*, 13. [\[CrossRef\]](#)
33. Yan, S.; Sui, L.; Wang, S.; Sun, Y. On-Line Tool Wear Monitoring under Variable Milling Conditions Based on a Condition-Adaptive Hidden Semi-Markov Model (CAHSMM). *Mech. Syst. Signal Process.* **2023**, *200*, 110644. [\[CrossRef\]](#)
34. Sharma, P.; Thulasi, H.M.; Mishra, S.K.; Ramkumar, J. Identification of Parameter-Dependent Machine Learning Models for Tool Flank Wear Prediction in Dry Titanium Machining. *Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng.* **2024**, 09544089241304236. [\[CrossRef\]](#)
35. Shurrab, S.; Almshnanah, A.; Duwairi, R. Tool Wear Prediction in Computer Numerical Control Milling Operations via Machine Learning. In *Proceedings of the 2021 12th International Conference on Information and Communication Systems (ICICS), Valencia, Spain, 24 May 2021*; IEEE: New York, NY, USA, 2021; pp. 220–227.
36. Misal, A.; Karandikar, H.; Sayyad, S.; Bongale, A.; Kumar, S.; Warke, V. Milling Tool Wear Estimation Using Machine Learning with Feature Extraction Approach. In *Proceedings of the 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSOciCon), Pune, India, 25 April 2024*; IEEE: New York, NY, USA, 2024; pp. 1–7.
37. Rather, I.H.; Kumar, S.; Gandomi, A.H. Breaking the Data Barrier: A Review of Deep Learning Techniques for Democratizing AI with Small Datasets. *Artif. Intell. Rev.* **2024**, *57*, 226. [\[CrossRef\]](#)
38. Danish, M.; Gupta, M.K.; Irfan, S.A.; Ghazali, S.M.; Rathore, M.F.; Krolczyk, G.M.; Alsaady, A. Machine Learning Models for Prediction and Classification of Tool Wear in Sustainable Milling of Additively Manufactured 316 Stainless Steel. *Results Eng.* **2024**, *22*, 102015. [\[CrossRef\]](#)
39. Dilli Ganesh, V.; Thangaraj, S.J.J. Prediction of Flank Wear in Turning of Monel K500 by Using Machine Learning Model in Comparison With Experimental Analysis. In *Proceedings of the 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), Chennai, India, 21 December 2023*; IEEE: New York, NY, USA, 2023; pp. 1–5.
40. Kosarac, A.; Mladjenovic, C.; Zeljkovic, M.; Tabakovic, S.; Knezev, M. Neural-Network-Based Approaches for Optimization of Machining Parameters Using Small Dataset. *Materials* **2022**, *15*, 700. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Zhu, Y.; Yuan, Z.; Khonsari, M.M.; Zhao, S.; Yang, H. Small-Dataset Machine Learning for Wear Prediction of Laser Powder Bed Fusion Fabricated Steel. *J. Tribol.* **2023**, *145*, 091101. [\[CrossRef\]](#)
42. Shah, R.; Pai, N.; Thomas, G.; Jha, S.; Mittal, V.; Shirvni, K.; Liang, H. Machine Learning in Wear Prediction. *J. Tribol.* **2025**, *147*, 040801. [\[CrossRef\]](#)
43. Hirsch, E.; Friedrich, C. Data-Driven Tool Wear Prediction in Milling, Based on a Process-Integrated Single-Sensor Approach. *arXiv* **2024**, arXiv:2412.19950.
44. Dubey, V.; Sharma, A.K.; Pimenov, D.Y. Prediction of Surface Roughness Using Machine Learning Approach in MQL Turning of AISI 304 Steel by Varying Nanoparticle Size in the Cutting Fluid. *Lubricants* **2022**, *10*, 81. [\[CrossRef\]](#)
45. Liu, Z.; Xu, Y.; Qiu, C.; Tan, J. A Novel Support Vector Regression Algorithm Incorporated with Prior Knowledge and Error Compensation for Small Datasets. *Neural Comput. Appl.* **2019**, *31*, 4849–4864. [\[CrossRef\]](#)
46. Schulz, E.; Speekenbrink, M.; Krause, A. A Tutorial on Gaussian Process Regression: Modelling, Exploring, and Exploiting Functions. *J. Math. Psychol.* **2018**, *85*, 1–16. [\[CrossRef\]](#)
47. Norazman, S.H.; Aspar, M.A.S.M.; Ghafar, A.N.A.; Karumdin, N.; Abidin, A.N.S.Z. Artificial Neural Network Analysis in Road Crash Data: A Review on Its Potential Application in Autonomous Vehicles. In *Intelligent Manufacturing and Mechatronics*; Isa, W.H.M., Khairuddin, I.M., Razman, M.A.M., Saruchi, S.A., Teh, S.-H., Liu, P., Eds.; Lecture Notes in Networks and Systems; Springer Nature: Singapore, 2024; Volume 850, pp. 95–104.
48. Tazi, K.; Lin, J.A.; Viljoen, R.; Gardner, A.; John, S.; Ge, H.; Turner, R.E. Beyond Intuition, a Framework for Applying GPs to Real-World Data. *arXiv* **2023**, arXiv:2307.03093. [\[CrossRef\]](#)
49. Ougiaroglou, S.; Evangelidis, G. Dealing with Noisy Data in the Context of K-NN Classification. In *Proceedings of the 7th Balkan Conference on Informatics Conference, Craiova, Romania, 2 September 2015*; ACM: New York, NY, USA, 2015; pp. 1–4.
50. Chaudhuri, A. Hierarchical Modified Regularized Least Squares Fuzzy Support Vector Regression through Multiscale Approach. In *Advances in Computational Intelligence*; Rojas, I., Joya, G., Gabestany, J., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7902, pp. 393–407, ISBN 978-3-642-38678-7.
51. Acito, F. K Nearest Neighbors. In *Predictive Analytics with KNIME*; Springer Nature: Cham, Switzerland, 2023; pp. 209–227, ISBN 978-3-031-45629-9.

52. Shrivastava, A.; Kotiyal, A.; Habelalmateen, M.I.; Rana, A.; Devi, V.S.A.; Rao, B.D.; Bansal, S. Leveraging XGBoost for Predictive Analytics in Healthcare: Enhancing Disease Diagnosis. In *Proceedings of the 2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, Greater Noida, India, 18 September 2024; IEEE: New York, NY, USA, 2024; pp. 1666–1672.
53. Yan, R.; Wang, S. Linear Regression Models. In *Applications of Machine Learning and Data Analytics Models in Maritime Transportation*; Institution of Engineering and Technology: London, UK, 2022; pp. 51–62, ISBN 978-1-83953-559-8.
54. Wan, A.; Gong, Z.; Chen, T.; AL-Bukhaiti, K. Mass Flow Characteristics Prediction of Refrigerants through Electronic Expansion Valve Based on XGBoost. *Int. J. Refrig.* **2024**, *158*, 345–352. [\[CrossRef\]](#)
55. Kretowski, M. Oblique and Mixed Decision Trees. In *Evolutionary Decision Trees in Large-Scale Data Mining: Studies in Big Data*; Springer International Publishing: Cham, Switzerland, 2019; Volume 59, pp. 101–113, ISBN 978-3-030-21850-8.
56. Chen, Y.; Dong, Y.; Liu, W. Prediction of Credit Default Based on the XGBoost Model. *Appl. Comput. Eng.* **2024**, *96*, 85–92. [\[CrossRef\]](#)
57. Lin, Z.; Fan, Y.; Tan, J.; Li, Z.; Yang, P.; Wang, H.; Duan, W. Tool wear prediction based on XGBoost feature selection combined with PSO-BP network. *Sci. Rep.* **2025**, *15*, 3096. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Qi, Y. Random Forest for Bioinformatics. In *Ensemble Machine Learning*; Zhang, C., Ma, Y., Eds.; Springer: New York, NY, USA, 2012; pp. 307–323, ISBN 978-1-4419-9325-0.
59. Wang, L.; Li, Q.; Yu, Y.; Liu, J. Region Compatibility Based Stability Assessment for Decision Trees. *Expert Syst. Appl.* **2018**, *105*, 112–128. [\[CrossRef\]](#)
60. Utkin, L.V.; Kovalev, M.S.; Frank Coolen, P.A. Robust Regression Random Forests by Small and Noisy Training Data. In *Proceedings of the 2019 XXII International Conference on Soft Computing and Measurements (SCM)*, St. Petersburg, Russia, 23–25 May 2019; IEEE: New York, NY, USA, 2019; pp. 134–137.
61. Pukelis, L.; Stančiauskas, V. The Opportunities and Limitations of Using Artificial Neural Networks in Social Science Research. *Politologija* **2019**, *94*, 56–80. [\[CrossRef\]](#)
62. Kim, D.; Lee, C.; Hwang, S.; Jeong, M.K. A Robust Support Vector Regression with a Linear-Log Concave Loss Function. *J. Oper. Res. Soc.* **2016**, *67*, 735–742. [\[CrossRef\]](#)
63. McKearnan, S.B.; Vock, D.M.; Marai, G.E.; Canahuate, G.; Fuller, C.D.; Wolfson, J. Feature Selection for Support Vector Regression Using a Genetic Algorithm. *Biostatistics* **2023**, *24*, 295–308. [\[CrossRef\]](#)
64. Liu, B.; Gao, L.; Li, B.; Marcos-Martinez, R.; Bryan, B.A. Nonparametric Machine Learning for Mapping Forest Cover and Exploring Influential Factors. *Landsc. Ecol.* **2020**, *35*, 1683–1699. [\[CrossRef\]](#)
65. Scornet, E. Random Forests and Kernel Methods. *IEEE Trans. Inf. Theory* **2016**, *62*, 1485–1500. [\[CrossRef\]](#)
66. Boldini, D.; Grisoni, F.; Kuhn, D.; Friedrich, L.; Sieber, S.A. Practical Guidelines for the Use of Gradient Boosting for Molecular Property Prediction. *J. Cheminform.* **2023**, *15*, 73. [\[CrossRef\]](#)
67. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A Comparative Analysis of Gradient Boosting Algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [\[CrossRef\]](#)
68. Che, Z.; Peng, C.; Wang, C.; Wang, J. A Novel Integrated TDLAVOA-XGBoost Model for Tool Wear Prediction in Lathe and Milling Operations. *Results Eng.* **2025**, *27*, 105984. [\[CrossRef\]](#)
69. Patra, K.; Jha, A.K.; Szalay, T.; Ranjan, J.; Monostori, L. Artificial Neural Network Based Tool Condition Monitoring in Micro Mechanical Peck Drilling Using Thrust Force Signals. *Precis. Eng.* **2017**, *48*, 279–291. [\[CrossRef\]](#)
70. Shahinur, S.; Ullah, A.M.M.S.; Noor-E-Alam, M.; Haniu, H.; Kubo, A. A Decision Model for Making Decisions under Epistemic Uncertainty and Its Application to Select Materials. *Artif. Intell. Eng. Des. Anal. Manuf.* **2017**, *31*, 298–312. [\[CrossRef\]](#)
71. ISO 3685; Tool-Life Testing with Single-Point Turning Tools. ISO: Geneva, Switzerland, 1993.
72. Li, G.; Li, N.; Wen, C.; Ding, S. Investigation and Modeling of Flank Wear Process of Different PCD Tools in Cutting Titanium Alloy Ti6Al4V. *Int. J. Adv. Manuf. Technol.* **2018**, *95*, 719–733. [\[CrossRef\]](#)
73. Fang, N.; Pai, P.S.; Mosquea, S. The Effect of Built-up Edge on the Cutting Vibrations in Machining 2024-T351 Aluminum Alloy. *Int. J. Adv. Manuf. Technol.* **2010**, *49*, 63–71. [\[CrossRef\]](#)
74. Kovvuri, V.; Wang, Z.; Araujo, A.; Da Silva, M.B.; Bukkapatnam, S.; Hung, W.N.P. Built-Up-Edge Formation in Micromilling. In *Proceedings of the Volume 2A: Advanced Manufacturing*, Houston, TX, USA, 13 November 2015; American Society of Mechanical Engineers: Houston, TX, USA, 2015; p. V02AT02A057.
75. Oliaei, S.N.B.; Karpas, Y. Investigating the Influence of Friction Conditions on Finite Element Simulation of Microscale Machining with the Presence of Built-up Edge. *Int. J. Adv. Manuf. Technol.* **2017**, *90*, 819–829. [\[CrossRef\]](#)
76. Faculty of Engineering; Thammasat School of Engineering (TSE); Thammasat University; Winnuwat, N.; Muttamara, A.; Kloypayan, J. A study of the phenomenon bue creation in trochoidal milling. *MM Sci. J.* **2023**, *2023*, 6435–6440. [\[CrossRef\]](#)
77. Sadeghi, M.; Behnia, F.; Amiri, R. Window Selection of the Savitzky–Golay Filters for Signal Recovery From Noisy Measurements. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 5418–5427. [\[CrossRef\]](#)
78. Krishnan, S.R.; Seelamantula, C.S. On the Selection of Optimum Savitzky–Golay Filters. *IEEE Trans. Signal Process.* **2013**, *61*, 380–391. [\[CrossRef\]](#)

79. Kondo, E.; Kamo, R.; Murakami, H. Monitoring of Burr and Prefailure Phase Caused by Tool Wear in Micro-Drilling Operations Using Thrust Force Signals. *J. Adv. Mech. Des. Syst. Manuf.* **2012**, *6*, 885–897. [\[CrossRef\]](#)
80. Li, G.S.; Lau, W.S.; Zhang, Y.Z. In-Process Drill Wear and Breakage Monitoring for a Machining Centre Based on Cutting Force Parameters. *Int. J. Mach. Tools Manuf.* **1992**, *32*, 855–867. [\[CrossRef\]](#)
81. Wang, H.; Liang, Q.; Hancock, J.T.; Khoshgoftaar, T.M. Feature Selection Strategies: A Comparative Analysis of SHAP-Value and Importance-Based Methods. *J. Big Data* **2024**, *11*, 44. [\[CrossRef\]](#)
82. Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874. [\[CrossRef\]](#)
83. Schumde, P. Feature Selection in Multiple Linear Regression Problems with Fewer Samples Than Features. In *Bioinformatics and Biomedical Engineering*; Rojas, I., Ortuño, F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2017; Volume 10208, pp. 85–95, ISBN 978-3-319-56147-9.
84. Baumann, K. Cross-Validation as the Objective Function for Variable-Selection Techniques. *TrAC Trends Anal. Chem.* **2003**, *22*, 395–406. [\[CrossRef\]](#)
85. Qiu, J. An Analysis of Model Evaluation with Cross-Validation: Techniques, Applications, and Recent Advances. *Adv. Econ. Manag. Polit. Sci.* **2024**, *99*, 69–72. [\[CrossRef\]](#)
86. Tantithamthavorn, C.; McIntosh, S.; Hassan, A.E.; Matsumoto, K. An Empirical Comparison of Model Validation Techniques for Defect Prediction Models. *IEEE Trans. Softw. Eng.* **2017**, *43*, 1–18. [\[CrossRef\]](#)
87. Lumumba, V.; Kiprotich, D.; Mpaine, M.; Makena, N.; Kavita, M. Comparative Analysis of Cross-Validation Techniques: LOOCV, K-Folds Cross-Validation, and Repeated K-Folds Cross-Validation in Machine Learning Models. *Am. J. Theor. Appl. Stat.* **2024**, *13*, 127–137. [\[CrossRef\]](#)
88. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Aroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* **2017**, *40*, 913–929. [\[CrossRef\]](#)
89. Mohr, F.; Van Rijn, J.N. Fast and Informative Model Selection Using Learning Curve Cross-Validation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 9669–9680. [\[CrossRef\]](#) [\[PubMed\]](#)
90. Zhang, P. Model Selection Via Multifold Cross Validation. *Ann. Stat.* **1993**, *21*, 299–313. [\[CrossRef\]](#)
91. Feng, C.-X.J.; Yu, Z.-G.S.; Emanuel, J.T.; Li, P.-G.; Shao, X.-Y.; Wang, Z.-H. Threefold versus Fivefold Cross-Validation and Individual versus Average Data in Predictive Regression Modelling of Machining Experimental Data. *Int. J. Comput. Integr. Manuf.* **2008**, *21*, 702–714. [\[CrossRef\]](#)
92. Wainer, J.; Cawley, G. Nested Cross-Validation When Selecting Classifiers Is Overzealous for Most Practical Applications. *Expert Syst. Appl.* **2021**, *182*, 115222. [\[CrossRef\]](#)
93. Szeghalmy, S.; Fazekas, A. A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning. *Sensors* **2023**, *23*, 2333. [\[CrossRef\]](#)
94. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Bischl, B.; Binder, M.; Lang, M.; Pielok, T.; Richter, J.; Coors, S.; Thomas, J.; Ullmann, T.; Becker, M.; Boulesteix, A.; et al. Hyperparameter Optimization: Foundations, Algorithms, Best Practices, and Open Challenges. *WIREs Data Min. Knowl. Discov.* **2023**, *13*, e1484. [\[CrossRef\]](#)
96. Lu, Z.; Dai, Y.; Li, W.; Su, Z. Joint Data and Feature Augmentation for Self-Supervised Representation Learning on Point Clouds. *Graph. Models* **2023**, *129*, 101188. [\[CrossRef\]](#)
97. Liu, D.; Kababji, S.E.; Mitsakakis, N.; Pilgram, L.; Walters, T.; Clemons, M.; Pond, G.; El-Hussuna, A.; Emam, K.E. Synthetic Data Generation for Augmenting Small Samples. *arXiv* **2025**, arXiv:2501.18741. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.