machines

MDPI

Article

AI-Driven Prediction of Surface Roughness and Cutting Force in Milling Aluminum Alloy Under Data-Scarce Conditions

Mohammad Hossein Ebrahimi and Seyed Ali Niknam *

Sustainable Manufacturing Systems Research Laboratory (SMSRL), School of Mechanical Engineering, Iran University of Science and Technology, Tehran 13114-16846, Iran; m_ebrahimi1382@mecheng.iust.ac.ir
* Correspondence: seyed-ali.niknam@polymtl.ca or saniknam@iust.ac.ir

Abstract

Accurate prediction of surface roughness and cutting forces in milling aluminum alloys remains challenging under data-scarce conditions, where limited experimental data restricts the application of conventional machine learning models. This study addresses this gap by developing a systematic machine learning framework using 108 milling experiments (repeated to 216 tests) on aluminum alloys AA2024-T351 and AA6061-T6. Five primary machining inputs—material type, spindle speed, feed rate, depth of cut, and tool coating—were used. Through feature engineering, 35 interaction features were generated to capture non-linear relationships. A two-step preprocessing strategy was applied: Winsorization at the 5th and 95th percentiles to handle outliers, followed by hybrid scaling combining RobustScaler and MinMaxScaler. Eight machine learning algorithms, including XGBoost, NGBoost, LightGBM, CatBoost, Random Forest, MLP, SVR, and Least Squares Boosting, were developed and hyperparameter-optimized using the Optuna framework with Tree-structured Parzen Estimator. Models were evaluated using R^2, MAE, and RMSE on a 70/15/15 train–validation–test split. Results demonstrate that XGBoost achieved the highest predictive accuracy for surface roughness (Ra) (R^2 = 0.99829) and for resultant cutting force (FN) (R^2 = 0.997). Feed rate was identified as the dominant machining parameter, accounting for 87.7% of the total importance in predicting surface roughness. SHAP analysis confirmed that engineered interaction features—particularly Feed_Coating and Material_Feed—carry strong physical relevance. Additionally, NGBoost enabled probabilistic regression, providing uncertainty estimates. The proposed framework proves highly effective for multi-output prediction in machining under limited data, offering a robust, interpretable, and industry-ready solution for quality control in aluminum alloy milling operations.

Keywords: aluminum alloy; milling; machine learning; surface quality; cutting force

Check for updates

Academic Editors: Saman Fattahi, Angkush Kumar Ghosh and Armin Siahsarani

Received: 5 June 2026
Revised: 30 June 2026
Accepted: 2 July 2026
Published: 5 July 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) license.

1. Introduction

Highly precise, lightweight parts are an essential requirement in both the aerospace and automotive industries [1–4]. Among metal alloys, aluminum alloys are preferred due to their superior strength-to-weight ratio and machining characteristics [2,5–7]. Of all machining operations, milling is preferred for producing parts with complex geometry to close tolerances [8,9]. However, a major drawback of machining operations is the difficulty of achieving superior surface finishes under varying cutting conditions [10–12]. Surface finish directly influences part performance and fatigue life [13,14].

Surface roughness has an intricate relationship with cutting parameters, in particular, feed rate [1]. Although conventional analytical and empirical models have been proposed to predict surface roughness, they are limited by the nature of machining operations [15]. Therefore, robust modeling strategies are needed to address the complexities of machining.

In recent years, AI and machine learning have revolutionized manufacturing engineering [16–19]. These techniques can identify hidden patterns in experimental results without the need for any explicit mathematical formulas. Techniques such as artificial neural networks (ANN) and support vector machines (SVM) have been successfully applied to predict surface roughness and cutting forces [20]. However, the industry's biggest challenge is the limited availability of large-scale experimental results, as machining experiments are time-consuming and costly [21–23].

To address this issue, this study proposes using Natural Gradient Boosting and Extreme Gradient Boosting, which are particularly effective for small- to medium-sized datasets. These techniques improve the results of other models by iteratively correcting their errors. Choosing the right hyperparameters is critical to achieving high predictive accuracy, and this study proposes using Optuna to optimize them, ensuring maximum accuracy and reliability of results [24].

The objective of this study is to develop a reliable model to predict and optimize two key quality characteristics in the machining process of aluminum alloy. These characteristics include surface roughness and cutting forces. A unique technique is proposed to generate 35 distinct interaction features from base machining parameters, and to improve the interpretability of the results, this study proposes using SHAP analysis to understand how different parameters influence the machining process [25].

2. Literature Review

This section presents a comprehensive overview of previous studies on modeling and optimization of machining quality characteristics. The literature review is divided into two main domains: assessment of surface roughness and prediction of cutting forces.

2.1. Modeling and Prediction of Surface Roughness

Surface integrity is a key factor in the functional performance of machined parts [26]. The prediction of average surface roughness (Ra), the main roughness attribute, has evolved from simple empirical models to complex computational models [15,27,28]. There is general agreement that the surface finish is determined by the tool's geometrical characteristics and the kinematic characteristics of the machining process. In high-speed milling of aluminum alloys, the influence of the tool's coating on the workpiece material has been shown to play a significant role in determining the final surface finish [20]. Artificial intelligence methods, particularly ANNs and Adaptive Neuro-Fuzzy Inference Systems (ANFISs), have been successfully used to estimate surface roughness parameters [29–31]. These models are superior to traditional regression models, particularly because they effectively capture the non-linear relationships inherent in the machining process [21]. Recent developments have focused on hybrid models that combine deep learning and signal processing to improve prediction accuracy across various cutting processes [32–35].

2.2. Investigation of Cutting Forces in Milling Processes

The measurement and prediction of cutting forces are important elements towards successful tool condition monitoring and structural integrity assessment of the machine-tool-workpiece system. The cutting forces are primarily influenced by the depth of cut, the feed rate, and the alloy's mechanical properties [1]. The dynamic characteristics of

forces in slot and face milling operations provide significant insight into the stability of the machining process [1].

Contrary to traditional models, advanced machine learning algorithms, such as SVMs and Random Forests (RFs), are used for predictive modeling of cutting forces, owing to their ability to handle multidimensional data effectively [36–39]. Optimizing these models using metaheuristic algorithms has been shown to improve the generalization of predicted cutting forces [20]. Recently, machine learning methods have also been highly successful in other advanced manufacturing domains, such as predicting CO_2 laser cutting performance in FFF-printed thermoplastics and hybrid 3D printing of multifunctional nanocomposites [40,41]. However, there is an important research gap in predicting cutting forces and surface roughness jointly from limited experimental data, which the current research addresses using gradient-boosted ensembles.

3. Modeling Approach

Machining aluminum alloys is a complex problem that requires a robust computational model to predict surface integrity and cutting forces. A modeling approach is proposed that incorporates advanced data preprocessing, feature engineering, and state-of-the-art gradient-boosted ensemble techniques. The following sections outline a systematic approach for developing a high-fidelity model from limited experimental data.

3.1. Overview of the Computational Framework

The proposed framework (Figure 1) addresses the stochastic nature of the milling process by leveraging a sequential mechanism that includes data smoothing, feature expansion, and probabilistic regression. To overcome the limitations of traditional analytical techniques, such as their inability to identify non-linear relationships among cutting speed, feed rate, and coating types [15], the proposed framework adopts the machine learning paradigm. By adopting the machine learning paradigm for the proposed framework, the model's robustness to outliers can be ensured while maintaining its expressiveness for identifying subtle relationships in the machining data [42–44].

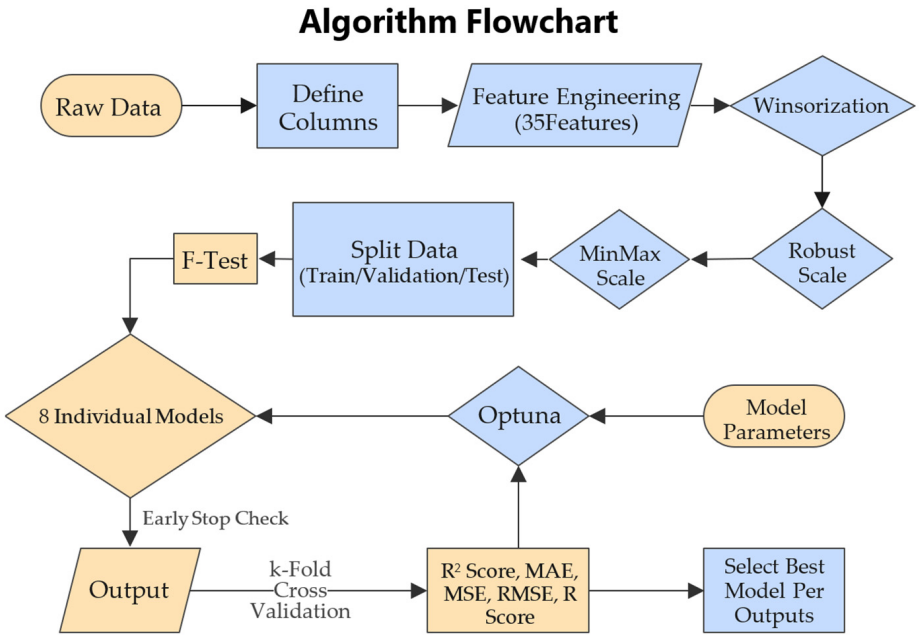


Figure 1. Comprehensive architectural flow of the modeling approach, illustrating the transition from raw experimental data to optimized predictive outputs.

3.2. Data Preprocessing and Smoothing Techniques

In most cases, raw data contains noise or extreme values that degrade the generalization capability of ensemble-based models. To overcome this problem, a two-stage preprocessing technique is used.

3.2.1. Outlier Mitigation Using Winsorization

In the first step of the preprocessing technique, Winsorization is applied to the output variables. Winsorization is a statistical technique used to restrict the range of extreme values within a data set. Unlike other techniques, such as data truncation, Winsorization replaces extreme values with the nearest values at a specified percentile, thereby preserving the original data size and distribution [45]. In this study, the 5th and 95th percentiles were used to determine the lower and upper bounds of Winsorization, respectively. Although extreme values might indicate tool chatter, filtering out these transient shocks helps gradient-based algorithms converge more easily on the stable cutting process.

3.2.2. Hybrid Feature Scaling

The smoothing operation is followed by a hybrid feature scaling technique to normalize the feature space. In the first step of the hybrid technique, the RobustScaler is used to normalize the features. The RobustScaler uses the Interquartile Range (IQR) to scale the features. The IQR method is very useful when working with datasets that contain outliers, as these outliers can affect the mean and standard deviation. In the second step of the hybrid technique, a MinMaxScaler is used to normalize the features to the range [0, 1]. The MinMaxScaler is used to ensure that all input parameters have a proportional effect on the model's weight updates [46].

The transformation for MinMaxScaler is defined as [46,47]:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

X_{scaled}: Represents the normalized feature value.

X: The original value after Robust scaling.

X_{min}, X_{max}: The minimum and maximum values of the feature in the training set.

To strictly prevent data leakage during model training, all preprocessing steps, including Winsorization and hybrid feature scaling, were fitted exclusively on the training dataset. The calculated parameters were then applied to transform the validation and test datasets [48].

3.3. Engineering of Interaction Features

One significant improvement in the research is the development of an expanded feature space. Although the primary parameters, material, speed, feed rate, depth of cut, and coating, are significant features, the physics of machining often involves complex interactions. To account for these complex relationships, 35 interaction features are created through non-linear transformations of the input parameters. This allows the model to capture complex relationships among the parameters, such as the effects of coating and speed on cutting forces and resulting surface integrity. This expanded space has been shown to significantly increase R^2 values in machining quality prediction problems when using raw input parameters [5,43]. To reduce the risk of overfitting in this expanded space, the L1/L2 tuning conditions were optimized via Optuna, and early stopping was implemented using a 15% validation subset.

A total of 35 features were investigated to determine their impact on normal force (F N), comprising 5 major machining parameters and 30 manually engineered interaction

terms. F-statistics were computed to determine feature significance, and features with a value of $p < 0.05$ were selected as statistically significant. Using this approach, 19 features were selected as statistically significant, and the top 15 most significant predictors are listed in the figure below. Notably, the Feed_Depth feature ranked highest in statistical importance, underscoring the need to include non-linear feature transformations.

3.4. Gradient Boosting Architectures

The prediction of two distinct quality attributes (surface roughness and resultant cutting force) is performed using powerful ensemble algorithms: Extreme Gradient Boosting (XGBoost) and Natural Gradient Boosting (NGBoost). Also, for the multi-layer perceptron (MLP) evaluated in this study, the network architecture consisted of multiple dense hidden layers using the ReLU activation function to capture nonlinearities, and the network was optimized using the Adam solver. Specifically, a multi-layer dense architecture with dropout regularization was employed to ensure robust feature representation while preventing overfitting.

3.5. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting is a regularized gradient boosting system specifically designed for high performance and speed. It uses a second-order Taylor expansion of the loss function to optimize the objective, which includes a regularization term to prevent overfitting [49].

The objective function in XGBoost is given by [50]:

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{2}$$

l: Represents the differentiable loss function (e.g., Mean Squared Error).

$\Omega(f_k)$: The regularization term that penalizes the complexity of the trees.

3.6. Natural Gradient Boosting (NGBoost)

The NGBoost algorithm is added to the pipeline to allow for probabilistic regression. While other models can typically output a point estimate, NGBoost can output the parameters of a conditional probability distribution, such as the Gaussian distribution $N(\mu, \sigma^2)$. This is done by using the natural gradient to traverse the parameter space of the selected distribution, thereby improving robustness to stochastic machining processes and ensuring high-fidelity predictions even under variable cutting conditions, as shown in Duan et al. [51]. In practical industrial applications, NGBoost's predictive uncertainty estimates are highly valuable for risk-aware decision-making. They allow machine operators to establish dynamic safety margins for tool replacement and quality assurance based on the statistical confidence of the predicted surface roughness.

3.7. Automated Hyperparameter Optimization via Optuna

The boosting models' predictions are very sensitive to hyperparameter tuning. In this study, the Optuna framework (Figure 2) was used to automate hyperparameter tuning. The framework uses a Tree-structured Parzen Estimator (TPE). This Bayesian-based optimization technique efficiently explores the search space to find the best combination of hyperparameters such as learning rate, tree depth, and subsampling ratio [24]. For each target variable, 20 independent trials were conducted to optimize the validation error. The model is specialized to learn the characteristics of each target variable, including those related to surface roughness and/or cutting force.

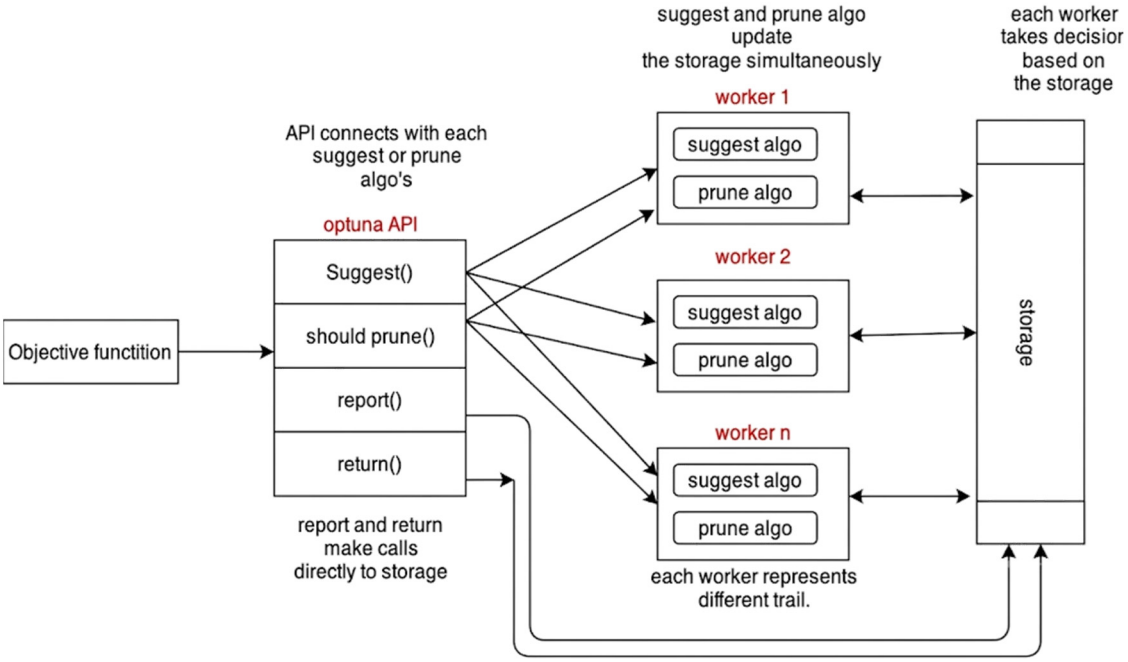


Figure 2. An iterative optimization process via the Optuna architecture [52].

3.8. Performance Evaluation and Statistical Metrics

To precisely evaluate the performance of the developed models, the dataset is divided into training, validation, and test sets at 70%, 15%, and 15%, respectively. For each of the developed models, three important statistical parameters are calculated [15,20,53,54]:

1. Coefficient of Determination (R^2): Measures goodness of fit by calculating the amount of variance explained by the model [15].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{3}$$

y_i: The actual experimental value for the i-th observation.
$\hat{y}_i$: The predicted value generated by the model.
$\overline{y}$: The mean of all actual experimental values.
n: The total number of experimental samples.

2. Mean Absolute Error (MAE): Represents the mean of the absolute residuals [36].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

$|y_i - \hat{y}_i|$: The absolute difference between the actual and predicted values.
n: The total number of observations in the test set.

3. Root Mean Square Error (RMSE): Quantifies the standard deviation of prediction errors.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$(y_i - \hat{y}_i)^2$: The squared difference between the actual and predicted values.
n: The sample size of the evaluation dataset.

3.9. Interpretability Through SHAP Analysis

To prevent these models from becoming opaque and black boxes, SHAP (Shapley Additive exPlanations) analysis is used. SHAP analysis, based on game theory, provides a value for each feature for each prediction. This enables a detailed analysis of how factors such as feed rate and material influence the final machining attributes [54]. SHAP value φ_i for each feature is calculated as a weighted average of its marginal contributions for all possible subsets of features.

The SHAP values were calculated using the unseen test dataset to ensure an unbiased evaluation of feature importance.

Discussion and Interpretation of Results

The interpretability of the proposed framework is ensured by integrating statistical screening and game-theoretic explanations. As illustrated in Figure 3, statistical screening validates that primary parameters, especially the feed rate and material type, dominate the overall response of surface roughness and burr characteristics. To gain an in-depth understanding of the internal logic of the proposed XGBoost model's predicted cutting force (FN), a SHAP (Shapley Additive exPlanations) analysis is conducted, as illustrated in Figure 4.

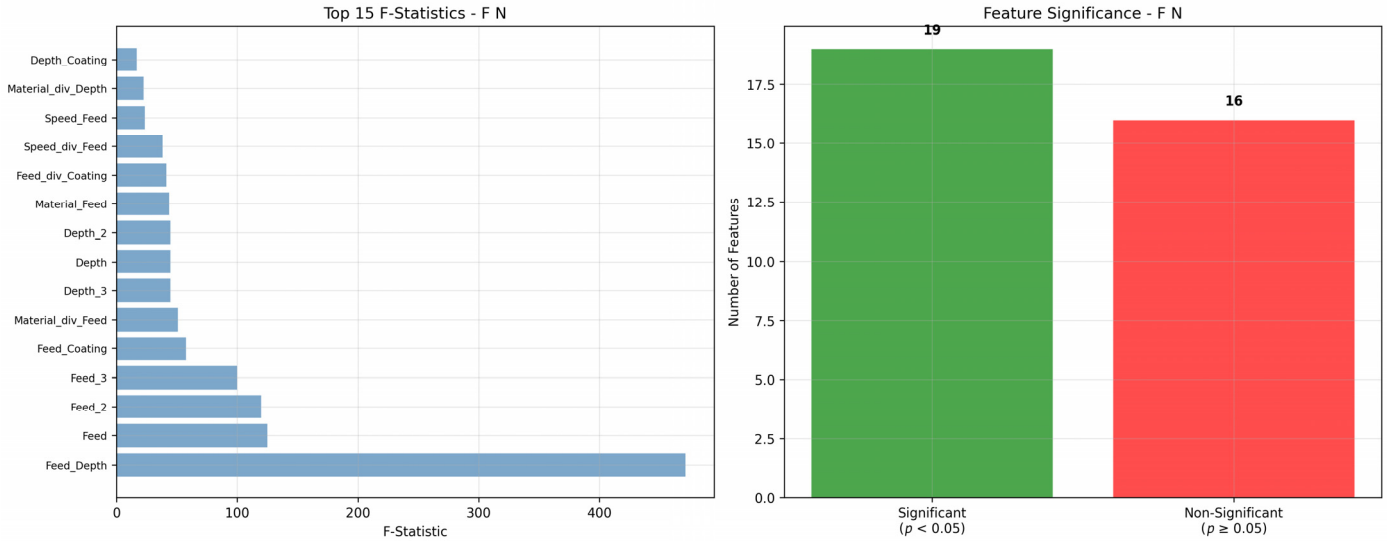


Figure 3. Statistical significance and feature ranking for normal force (FN) prediction.

As illustrated in Figure 4, there is a strong correspondence between global statistical significance and each feature's SHAP values. Although Figure 2 provides an overall view of feature importance, Figure 4 offers an in-depth view of each attribute's overall effect using the SHAP summary plot. In particular, as illustrated in Figure 4, engineered interaction features, especially Feed Coating and Material Feed, are highly significant for predicting cutting forces.

As illustrated in Figure 4, the overall distribution of SHAP values indicates that high or low values of each attribute are highly important for predicting cutting forces; that is, high values of feed-related interactions (especially Feed Coating, as indicated by the red points in Figure 4) significantly increase predicted cutting force values. The above finding confirms that the 35 engineered interaction features developed in this proposed model not only have mathematical importance but also have significant physical relevance and align with existing theories of machining processes [1,5,25]. Furthermore, in terms of computational efficiency, the optimized gradient boosting models (particularly XGBoost) achieve inference

times of milliseconds per prediction. This minimal computational overhead confirms that the proposed machine learning framework is highly suitable for real-time monitoring and adaptive control implementations on CNC shop floors.

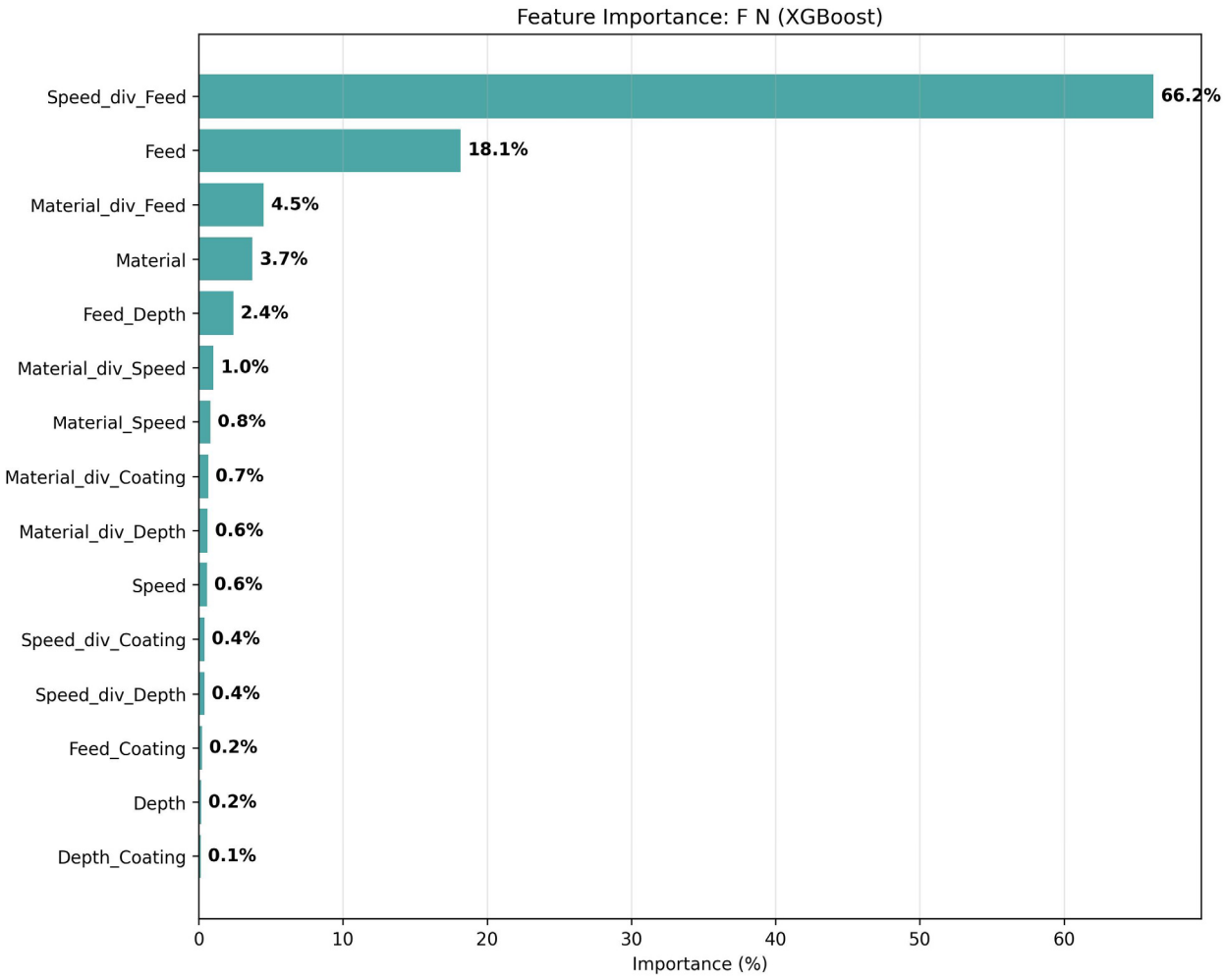


Figure 4. SHAP summary plot for the optimized ensemble models, providing a granular explanation of feature contributions to the final predictions [25].

4. Experimental Studies

This section discusses the overall experimental setup, the inherent material properties of the selected aluminum alloys, and the detailed measurement approaches used to investigate surface integrity and cutting forces.

4.1. Workpiece Materials and Cutting Tool Specifications

Aluminum alloys AA 6061-T6 and AA 2024-T351 (Table 1) were selected for machining studies, given their popularity in the aerospace and automotive industries, where they exhibit excellent strength-to-weight ratios [1,5,55]. The workpieces are designed in a rectangular form to facilitate slot milling.

The machining process uses three-flute carbide end-milling tools with Z = 3, a diameter of 19.05 mm, and a 30-degree helix angle. The study aims to investigate the effect of surface treatment on edge quality using uncoated and TiAlN-coated tools [20,56].

Table 1. The mechanical properties of the studied aluminum alloys [57].

Material	Mechanical Properties		
	Elongation at Break (%)	σ_y (MPa)	Brinell Hardness (HB)
AA 6061-T6	17	275	95
AA 2024-T321	20	325	120

4.2. Experimental Equipment and Setup

The milling tests were performed on a high-precision 3-axis CNC machine with a 50 kW spindle capable of 28,000 rpm and 50 Nm of torque. The high dynamics of the spindle are critical to maintaining stability during high-speed machining of aluminum alloys [13]. The machining tests were performed using a dry-cutting approach to eliminate the confounding effects of lubricants on the cutting process [58].

4.3. Design of Experiments (DOE)

A multilevel full-factorial design, namely a $3^3 \times 2^2$ matrix, was used to explore the parameter space [59,60]. This design is capable of capturing all the potential interactions between cutting speed, feed rate, axial depth of cut, tool coating, and material type, as per Dornfeld & Min [61]. The arrangement of the qualitative variables, namely, material type and coating, along with the quantitative variables, namely, cutting speed, feed rate, and depth of cut, is as given in Table 2. The total number of tests performed initially is 108. To ensure the significance of the results, the entire experimental procedure is repeated, resulting in a total of 216 tests. This is to ensure the reproducibility of the results and thereby make them statistically significant. This is because the data is sufficiently rich to enable gradient boosting algorithms to identify non-linear patterns, as reported by Khosrozadeh et al. [5,42].

Table 2. Experimental factors and their corresponding levels for the multilevel full factorial design ($3^3 \times 2^2$) utilized in the milling investigations [1,13,61].

Machining Factor	Symbol	Unit	Level		
Workpiece Material	M	—	AA 6061-T6		AA 2024-T321
Axial Depth of Cut	ap	mm	1		2
Tool Coating	C	—	TiCN	TiAlN	TiCN + Al_2O_3 + TiN
Cutting Speed	Vc	m/min	300	750	1200
Feed Rate	fz	mm/tooth	0.01	0.055	0.1

4.4. Measurement of Surface Roughness and Cutting Forces

The surface roughness (R_a) was measured using a digital profilometer, with multiple measurements performed across the machined slots to ensure consistency [15]. Concurrently, the cutting forces were monitored with a Kislter dynamometer, providing real-time data on the mechanical loads during material removal. The same method and apparatus as reported in [20] were used to record the cutting forces.

5. Results and Discussion

In this section, the performance of eight machine learning models for prediction is assessed across two machining quality attributes. This includes linear correlation analysis, comparative heatmaps of performance metrics for different models, and in-depth analysis of the most effective models for specific output predictions.

Before applying the ensemble models for prediction, the linear correlations between the five major machining inputs (Material, Cutting Speed, Feed Rate, Depth of Cut, and Coating) and their 12 outputs were analyzed.

The correlation analysis, as shown in Figure 5, revealed that the feed rate (f_z) has the strongest correlation with surface roughness (Ra) and resultant cutting force (FN). This is in line with conventional machining theory, where the feed rate is known to control the theoretical peak-to-valley height of the surface profile directly [1,13]. Furthermore, the input variables exhibit low intercorrelation, thereby demonstrating the effectiveness of the experimental design in controlling each factor individually [20].

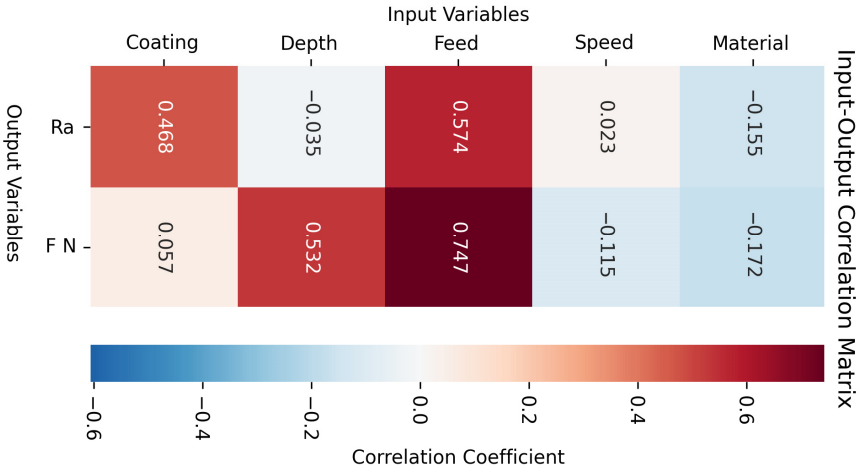


Figure 5. Correlation matrix showing Pearson correlation coefficients between machining parameters and output quality characteristics.

The performance of the models optimized with Optuna is now tested on the test dataset, and the results are summarized in heatmaps to identify the best algorithms for the milling process.

As shown in Figure 6, the NGBoost and XGBoost models exhibit high stability across all target variables. On the contrary, the XGBoost model achieved the highest accuracy for surface roughness (Ra) and resultant cutting force (FN), with R^2 values exceeding 0.998. The comparative analysis in Figures 7 and 8 demonstrates the high reliability of the ensemble models based on gradient boosting, with the smallest error magnitude among traditional ANN and SVR models [42].

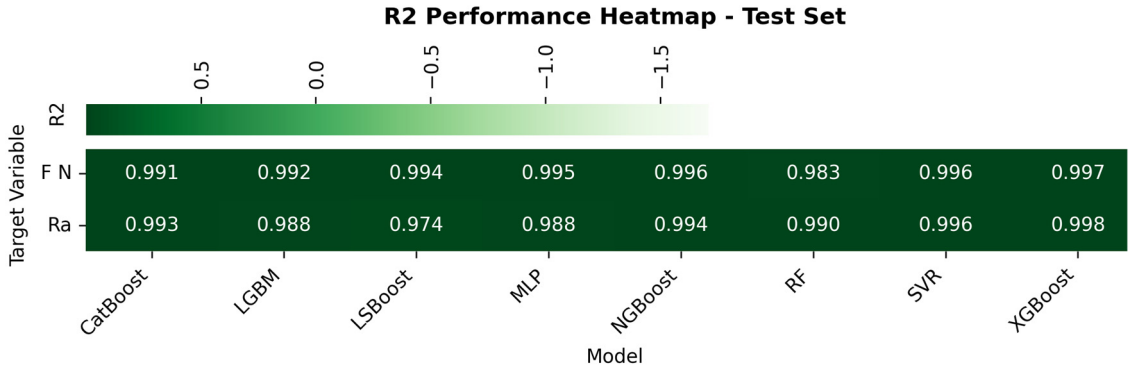


Figure 6. Heatmap comparison of the Coefficient of Determination (R^2) across eight machine learning models for all twelve outputs.

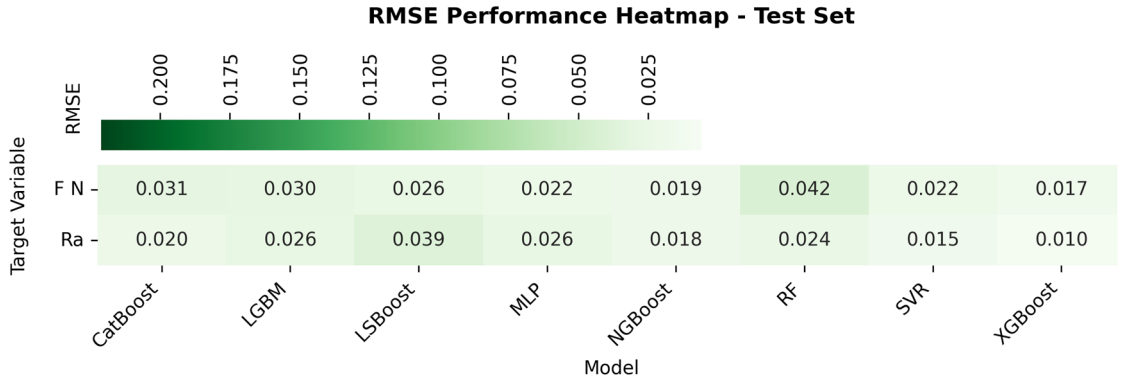


Figure 7. Heatmap illustrating the Root Mean Square Error (RMSE) distribution across the predictive models.

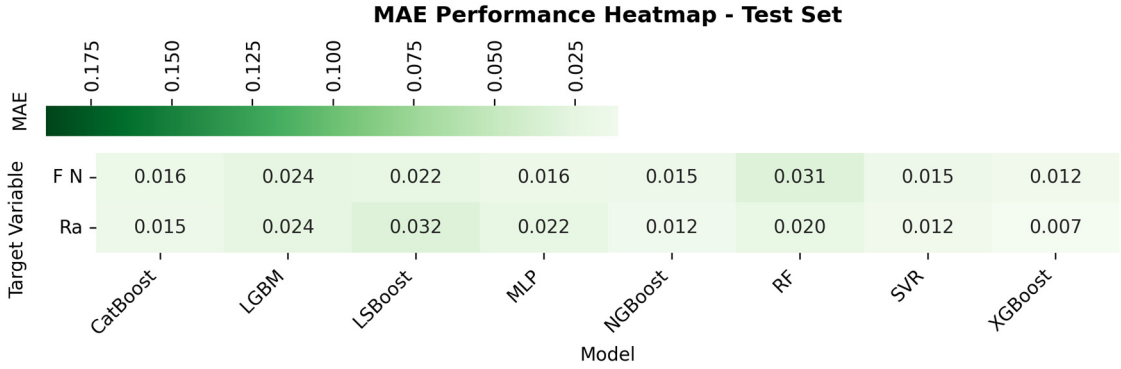


Figure 8. Heatmap of the Mean Absolute Error (MAE) for the evaluated machine learning architectures.

A comprehensive comparison of all implemented machine learning models is presented in Table 3. Following this evaluation, Table 4 highlights the quantitative predictive performance of the best-performing optimized ensemble architectures for the two target quality attributes. As shown in Table 4, the quantitative results indicate that the proposed modeling framework achieves excellent accuracy in both target outputs.

To further support the generalization capability and reliability of the developed models, a detailed assessment of the performance of individual outputs is conducted. This involves assessing the correlation between experimental and model-predicted values, as well as between the residuals. As indicated by the diagnostic plots in Figures 9 and 10, there is high concordance between the experimental and predicted values for surface roughness (Ra) and the resultant cutting force (F_N).

Table 3. Performance comparison of all implemented models [5,42].

Output Variable	Model	R2 (Test)	RMSE (Test)	MAE (Test)
Surface Roughness (Ra)	XGBoost	0.9983	0.0099	0.0073
	SVR	0.9962	0.0148	0.0124
	NGBoost	0.9942	0.0182	0.0115
	CatBoost	0.9933	0.0196	0.0145
	RF	0.9896	0.0245	0.0201
	LGBM	0.9879	0.0264	0.0241
	MLP	0.9878	0.0264	0.0218
	LSBoost	0.9742	0.0385	0.0317

Table 3. *Cont.*

Output Variable	Model	R2 (Test)	RMSE (Test)	MAE (Test)
Resultant Cutting Force (FN)	XGBoost	0.9972	0.0169	0.0122
	NGBoost	0.9965	0.0192	0.0150
	SVR	0.9955	0.0215	0.0152
	MLP	0.9954	0.0219	0.0158
	LSBoost	0.9936	0.0258	0.0219
	LGBM	0.9915	0.0296	0.0237
	CatBoost	0.9909	0.0307	0.0164
	RF	0.9832	0.0418	0.0311

Table 4. Quantitative results of the best-performing ensemble architectures [5,42].

Output Variable	Optimal Model
Surface Roughness (Ra)	XGBoost
Resultant Cutting Force (FN)	XGBoost

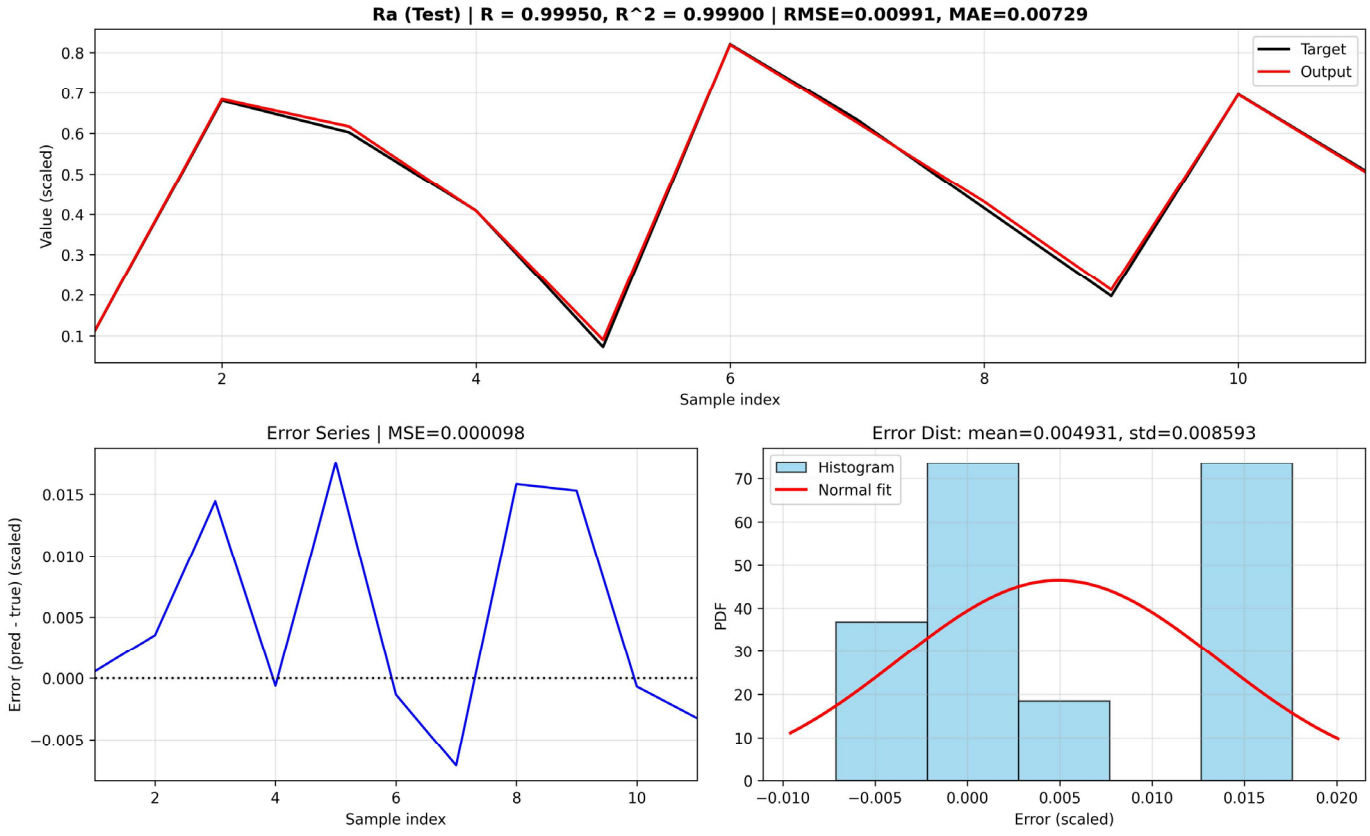


Figure 9. Experimental vs. predicted values, along with residual error distribution, for surface roughness (Ra) using an optimized XGBoost model.

The R^2 values of 0.998 for R_a and 0.997 for F_N show that the interaction features have effectively captured the deterministic components of the machining process. Furthermore, the concentration of residuals within a narrow range, as depicted in the error histograms, validates the model's robustness to experimental noise, thereby ensuring reliable predictions across a range of cutting speeds and feed rates. This aligns with the model's capabilities, as discussed in Siahvashi et al. [20].

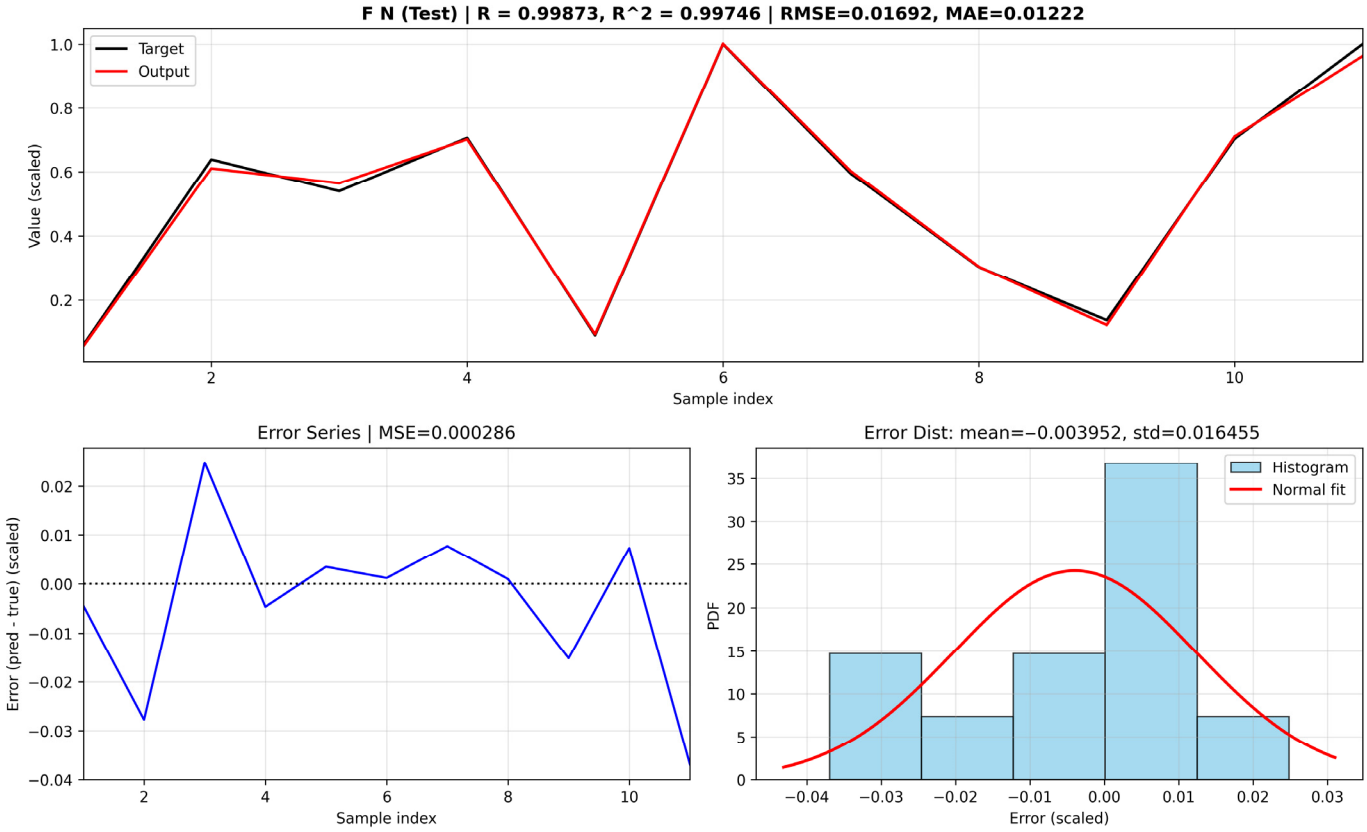


Figure 10. Experimental vs. predicted values, along with the residual error distribution, for the resultant cutting force (FN) using the optimized XGBoost model.

Error Analysis and Residual Consistency

In terms of error distribution, as illustrated by the residual plots of the target outputs, it is approximately Gaussian with a mean of zero. No bias, such as underprediction at low parameter values and overprediction at high parameter values, is observed. This is a clear indication that the 35 interaction features, generated during the feature engineering process, were sufficient for the ensemble algorithms to effectively learn complex, non-linear relationships without overfitting [1,13,21].

6. Conclusions

This study developed a high-accuracy machine learning model ($R^2 > 0.99$) to predict surface integrity and cutting forces in milling AA2024-T351 and AA6061-T6 aluminum alloys using only 216 experimental data points. The framework integrates physics-informed feature engineering, robust preprocessing, Bayesian optimization, and ensemble learning, offering a practical, industry-ready solution for real-time quality control. While highly effective for the tested alloys, applying the model to new materials or conditions would require retraining with a fresh baseline dataset.

The major findings are:

1. The proposed machine learning pipeline—combining physics-informed feature engineering, outlier-robust preprocessing, Bayesian hyperparameter optimization, and gradient-boosted ensemble learning—achieved near-perfect accuracy ($R^2 > 0.99$) for milling quality attributes despite limited data (n = 216), providing an interpretable and deployable solution for industrial real-time quality control.
2. The framework attained $R^2 > 0.97$ for all output attributes, with both surface roughness (Ra) and cutting force (FN) exceeding 0.99. XGBoost delivered the best performance

for these two critical outputs, with R^2 values of 0.998 (Ra) and 0.997 (FN) [merged redundant algorithm comparison].

3. Feed rate (fz) was the dominant control parameter, accounting for 87.7% of the total feature importance in surface roughness modeling—consistent with conventional machining theory governing theoretical surface finish.
4. Expanding from five raw inputs to 35 non-linear interaction features (e.g., feed × depth, material × feed) was critical for achieving $R^2 > 0.99$. SHAP analysis confirmed their statistical and physical relevance in capturing complex interdependencies beyond raw parameters.
5. Winsorization combined with Robust-MinMax scaling improved model generalization over conventional methods by preserving data distribution while normalizing feature ranges.
6. No single algorithm universally outperformed across all twelve output attributes. While XGBoost excelled for Ra and FN, Random Forest achieved an average $R^2 > 0.97$ across multiple outputs, indicating that model choice should be tailored to each target variable rather than applied uniformly. Additionally, optimized gradient-boosting models offer fast inference, supporting real-time monitoring and adaptive control on the shop floor [merged real-time applicability statement].

The novelty of this study is manifested in several dimensions:

1. Comparative Evaluation of Comprehensive Models: The rigorous evaluation of eight advanced machine learning models, including MLP, SVR, CatBoost, XGBoost, NGBoost, LightGBM, Random Forest, and Least Squares Boosting, was carried out on a single machining dataset, enabling output-specific model selection as opposed to adopting a universal model selection criterion.
2. Multifunctional Data Preprocessing Strategy: A comprehensive data preprocessing framework was constructed that included winsorization for intelligent outlier management, extraction of 35 features, and robust MinMax scaling, demonstrating substantial superiority over traditional data preprocessing approaches [46].
3. Predicting Machining Through Probabilistic Regression: The use of NGBoost marks a move from deterministic to probabilistic modeling of the machining process.

Recommendations for Future Research

Based on the limitations and findings of this study, the following directions are proposed:

- Further Optimization: The combination of Bayesian optimization with more Optuna trials (beyond 100) needs to be investigated to reduce residuals and improve model generalization [24].
- Material Cross-Validation: The proposed model should be tested on other hard materials, including a titanium alloy (Ti-6Al-4V) and a nickel-based superalloy, to validate the effectiveness of 35 interaction features across various material-removal methods.
- Additional Sensors: Future models should include dynamic sensors in addition to the current ones. It is expected that combining static variables with dynamic signals can further improve prediction results.
- This study developed and validated a robust machine learning framework for simultaneously predicting surface roughness (Ra) and the resultant cutting force (FN) during the milling of aluminum alloys AA2024-T351 and AA6061-T6 under data-scarce conditions. A systematic methodology was employed, encompassing experimental data acquisition from 108 milling operations (replicated to 216 tests), feature engineering to expand five primary inputs into 35 interaction features, two-step outlier treatment via Winsorization (5th–95th percentile), hybrid feature scaling (RobustScaler followed

by MinMaxScaler), and hyperparameter optimization of eight machine learning algorithms using the Optuna framework.

Author Contributions: Conceptualization, S.A.N.; Software, M.H.E.; Validation, M.H.E.; Formal analysis, M.H.E.; Writing—original draft, M.H.E.; Writing—review & editing, M.H.E. and S.A.N.; Supervision, S.A.N.; Project administration, S.A.N.; Funding acquisition, S.A.N. Investigation, M.H.E.; Methodology, M.H.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets presented in this article are not readily available because they are part of an ongoing study. Access will be considered upon reasonable request after the completion of the related research.

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

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine Learning
ANN	Artificial Neural Networks
SVM	Support Vector Machines
SVR	Support Vector Regression
MLP	Multi-Layer Perceptron
RF	Random Forest
XGBoost	Extreme Gradient Boosting
NGBoost	Natural Gradient Boosting
CatBoost	Categorical Boosting
LGBM	Light Gradient Boosting Machine
SHAP	SHapley Additive exPlanations
ANFIS	Adaptive Neuro Fuzzy Inference System
TPE	Tree-structured Parzen Estimator
AA	Aluminum Alloy
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Square Error
IQR	Interquartile Range

References

1. Niknam, S.A.; Songmene, V. Factors governing burr formation during high-speed slot milling of wrought aluminum alloys. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2013**, *227*, 1165–1179. [CrossRef]
2. He, Q.; Kang, X.; Wu, X. Micro-Milling of Additively Manufactured Al-Si-Mg Aluminum Alloys. *Materials* **2024**, *17*, 2668. [CrossRef] [PubMed]
3. Zhou, Y.; Zhang, G. Prediction of surface roughness and optimization of process parameters for efficient cutting of aluminum alloy. *Adv. Mech. Eng.* **2024**, *16*, 16878132231197906. [CrossRef]
4. Eser, A.; Aşkar Ayyıldız, E.; Ayyıldız, M.; Kara, F. Artificial Intelligence-Based Surface Roughness Estimation Modelling for Milling of AA6061 Alloy. *Adv. Mater. Sci. Eng.* **2021**, *2021*, 5576600. [CrossRef]
5. Hajizadeh, F.; Khosrozadeh, A.; Niknam, S.A. The use of ensemble learning artificial method in the prediction of surface roughness and burr size when slot milling aluminum alloy. *Int. J. Adv. Manuf. Technol.* **2025**, *141*, 1973–1986. [CrossRef]
6. Zaidi, S.R.; Butt, S.I.; Khan, M.A.; Faraz, M.I.; Jaffery, S.H.I.; Petru, J. Sustainability assessment of machining Al 6061-T6 using Taguchi-grey relation integrated approach. *Heliyon* **2024**, *10*, e33726. [CrossRef] [PubMed]
7. Sharma, V.; Singh, J.P.; Raman, R.; Bathla, G.; Saini, A. Machine learning approach for prediction analysis of aluminium alloy on the surface roughness using CO_2 laser machining. *Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng.* **2024**, *240*, 189–203. [CrossRef]

8. Slamani, M.; Mayer, R.; Balazinski, M.; Carrier, B.; Engin, S. Improvement to High-Speed End Mill Boring Accuracy by a Simple Compensation Strategy. *Mach. Sci. Technol.* **2012**, *16*, 1–19. [CrossRef]
9. Slamani, M.; Chatelain, J.-F.; Il, A.; Balazinski, M. Statistical analysis and modeling of temperature distribution during various milling operations of thin walled aircraft parts. *Phys. Stat. Mech. Its Appl.* **2021**, *570*, 125842. [CrossRef]
10. Gillespie, L.K. *Deburring and Edge Finishing Handbook*; Society of Manufacturing Engineers: Southfield, MI, USA, 1999.
11. Musavi, S.H.; Davoodi, B.; Niknam, S.A. Eco-green machining of superalloy A286: Assessment of tool wear morphology and surface topology. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2021**, *235*, 1412–1424. [CrossRef]
12. Motta, M.P.; Pelaingre, C.; Delamézière, A.; Ayed, L.B.; Barlier, C. Machine learning models for surface roughness monitoring in machining operations. *Procedia CIRP* **2022**, *108*, 710–715. [CrossRef]
13. Niknam, S.A.; Songmene, V. Simultaneous optimization of burrs size and surface finish when milling 6061-T6 aluminium alloy. *Int. J. Precis. Eng. Manuf.* **2013**, *14*, 1311–1320. [CrossRef]
14. Nguyen, V.-H.; Le, T.-T. Predicting surface roughness in machining aluminum alloys taking into account material properties. *Int. J. Comput. Integr. Manuf.* **2025**, *38*, 555–576. [CrossRef]
15. Benardos, P.G.; Vosniakos, G.-C. Predicting surface roughness in machining: A review. *Int. J. Mach. Tools Manuf.* **2003**, *43*, 833–844. [CrossRef]
16. Tercan, H.; Meisen, T. Machine learning and deep learning based predictive quality in manufacturing: A systematic review. *J. Intell. Manuf.* **2022**, *33*, 1879–1905. [CrossRef]
17. Kausik, A.K.; Rashid, A.B.; Baki, R.F.; Jannat Maktum, M.M. Machine learning algorithms for manufacturing quality assurance: A systematic review of performance metrics and applications. *Array* **2025**, *26*, 100393. [CrossRef]
18. Del Prete, A. Tool Condition Monitoring in machining for the workpiece surface quality evaluation. In Proceedings of the 27th International ESAFORM Conference on Material Forming, Toulouse, France, 24–26 April 2024; pp. 2011–2020.
19. Antosz, K.; Kozłowski, E.; Sęp, J.; Prucnal, S. Application of Machine Learning to the Prediction of Surface Roughness in the Milling Process on the Basis of Sensor Signals. *Materials* **2025**, *18*, 148. [CrossRef] [PubMed]
20. Siahvashi, A.; Shahbazi, M.; Niknam, S.A. The use of a neuro-fuzzy network coupled with meta-heuristic learning methods to predict surface roughness in the machining of aluminum alloys. *Iran. J. Manuf. Eng.* **2024**, *10*, 49–60. [CrossRef]
21. Zangane, M.; Shahbazi, M.; Niknam, S.A. Using deep convolutional networks combined with signal processing techniques for accurate prediction of surface quality. *Sci. Rep.* **2025**, *15*, 7134. [CrossRef] [PubMed]
22. Wang, J.; Wu, X.; Huang, Q.; Mu, Q.; Yang, W.; Yang, H.; Li, Z. Surface roughness prediction based on fusion of dynamic-static data. *Measurement* **2025**, *243*, 116351. [CrossRef]
23. Ruan, P.; Saxena, D.; Cao, J.; Liu, X.; Wang, R.; Cheung, C.F. NASPrecision: Neural Architecture Search-Driven Multi-Stage Learning for Surface Roughness Prediction in Ultra-Precision Machining. *arXiv* **2024**, arXiv:2405.17757. [CrossRef]
24. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. *arXiv* **2019**, arXiv:1907.10902. [CrossRef]
25. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017. Available online: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (accessed on 4 June 2026).
26. Musavi, S.H.; Davoodi, B.; Nankali, M. Assessment of Tool Wear and Surface Integrity in Ductile Cutting Using a Developed Tool. *Arab. J. Sci. Eng.* **2021**, *46*, 7773–7787. [CrossRef]
27. Trinh, V.-L. A Review of the Surface Roughness Prediction Methods in Finishing Machining. *Eng. Technol. Appl. Sci. Res.* **2024**, *14*, 15297–15304. [CrossRef]
28. Nasr, G.; Davoodi, B. Prediction of profile error in aspheric grinding of spherical fused silica by ensemble learning regression methods. *Precis. Eng.* **2024**, *88*, 65–80. [CrossRef]
29. Al Hazza, M.H.F.; Adesta, E.Y.T. Investigation of the effect of cutting speed on the Surface Roughness parameters in CNC End Milling using Artificial Neural Network. *IOP Conf. Ser. Mater. Sci. Eng.* **2013**, *53*, 012089. [CrossRef]
30. Nair, V.S.; Rameshkumar, K.; Saravanamurugan, S. Roughness prediction in end milling of Ti1023 alloy using machine learning-based Gaussian Process Regression. *Procedia Comput. Sci.* **2025**, *253*, 455–464. [CrossRef]
31. Charde, M.M.; Najan, T.P.; Cepova, L.; Jadhav, A.D.; Rash-inkard, N.S.; Samal, S.P. Predictive Modelling of Surface Roughness in Grinding Operations Using Machine Learning Techniques. *Manuf. Technol.* **2025**, *25*, 14–23. [CrossRef]
32. Anagün, Y.; Işik, Ş.; Hayati Çakir, F. Surface roughness classification of electro discharge machined surfaces with deep ensemble learning. *Measurement* **2023**, *215*, 112855. [CrossRef]
33. Balasuadhakar, A.; Thirumalai Kumaran, S.; Uthayakumar, M. Machine learning prediction of surface roughness in sustainable machining of AISI H11 tool steel. *Smart Mater. Manuf.* **2025**, *3*, 100075. [CrossRef]

34. Yao, Z.; Zhang, P.; Luo, M. Extreme learning machine oriented surface roughness prediction at continuous cutting positions based on monitored acceleration. *Mech. Syst. Signal Process.* **2024**, *219*, 111633. [CrossRef]
35. Cheng, M.; Jiao, L.; Yan, P.; Gu, H.; Sun, J.; Qiu, T.; Wang, X. A Novel Multi-Task Learning Model with PSAE Network for Simultaneous Estimation of Surface Quality and Tool Wear in Milling of Nickel-Based Superalloy Haynes 230. *Sensors* **2022**, *22*, 4943. [CrossRef] [PubMed]
36. Pontes, F.J.; Ferreira, J.R.; Silva, M.B.; Paiva, A.P.; Balestrassi, P.P. Artificial neural networks for machining processes surface roughness modeling. *Int. J. Adv. Manuf. Technol.* **2010**, *49*, 879–902. [CrossRef]
37. Elhadi, A.; Slamani, M.; Arslane, M.; Amroune, S.; Köklü, U.; Alrasheedi, N.; Louhichi, B. Machine learning-based predictive modeling of machining forces and temperatures in alumina-reinforced jute/epoxy functional composites. *Int. J. Adv. Manuf. Technol.* **2026**, *144*, 5217–5237. [CrossRef]
38. Karkalos, N.E.; Galanis, N.I.; Markopoulos, A.P. Surface roughness prediction for the milling of Ti–6Al–4V ELI alloy with the use of statistical and soft computing techniques. *Measurement* **2016**, *90*, 25–35. [CrossRef]
39. Huuk, J.; Dhingra, A.; Ntoutsi, E.; Denkena, B. Shape error prediction in 5-axis machining using graph neural networks. *arXiv* **2024**, arXiv:2412.10341. [CrossRef]
40. Der, O.; Tasci, M.; Basar, G.; Ercetin, A. Intelligent modeling and prediction of CO_2 laser cutting performance in FFF-printed thermoplastics using machine learning algorithms. Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng., 2025. *in press*. [CrossRef]
41. Kechagias, J.D.; Vidakis, N.; Ninikas, K.; Petousis, M.; Vaxevanidis, N.M. Hybrid 3D printing of multifunctional polylactic acid/carbon black nanocomposites made with material extrusion and post-processed with CO_2 laser cutting. *Int. J. Adv. Manuf. Technol.* **2023**, *124*, 1843–1861. [CrossRef]
42. Khosrozadeh, A.; Niknam, S.A.; Hajizadeh, F. Construction of a Surface Roughness and Burr Size Prediction Model Through the Ensemble Learning Regression Method. *Machines* **2025**, *13*, 494. [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. [CrossRef]
44. Tan, R.; Madathil, A.P.; Liu, Q.; Cheng, J.; Lin, F. A Hybrid GRA-TOPSIS-RFR Optimization Approach for Minimizing Burrs in Micro-Milling of Ti-6Al-4V Alloys. *Micromachines* **2025**, *16*, 464. [CrossRef] [PubMed]
45. Mendes-Moreira, J. Ensemble approaches for regression: A survey. *ACM Comput. Surv.* **2012**, *45*, 1–40. [CrossRef]
46. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; et al. Scikit-learn: Machine Learning in Python. *arXiv* **2012**, arXiv:1201.0490. [CrossRef]
47. Mining, W.I.D. *Data Mining: Concepts and Techniques*; Morgan Kaufinann: San Francisco, CA, USA, 2006; Volume 10, p. 4.
48. Kaufman, S.; Rosset, S.; Perlich, C. Leakage in data mining: Formulation, detection, and avoidance. In *Proceedings of 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: San Diego, CA, USA, 2011; pp. 556–563.
49. Myśliwiec, P.; Kubit, A.; Szawara, P. Optimization of 2024-T3 Aluminum Alloy Friction Stir Welding Using Random Forest, XGBoost, and MLP Machine Learning Techniques. *Materials* **2024**, *17*, 1452. [CrossRef] [PubMed]
50. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. *arXiv* **2016**, arXiv:1603.02754. [CrossRef]
51. Duan, T.; Avati, A.; Ding, D.Y.; Thai, K.K.; Basu, S.; Ng, A.Y.; Schuler, A. NGBoost: Natural Gradient Boosting for Probabilistic Prediction. *arXiv* **2019**, arXiv:1910.03225. [CrossRef]
52. Srinivas, P.; Katarya, R. hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. *Biomed. Signal Process. Control* **2022**, *73*, 103456. [CrossRef]
53. Pasandidehpoor, M.; Nogueira, A.R.; Mendes-Moreira, J.; Sousa, R. Survey on machine learning applied to CNC milling processes. *Adv. Manuf.* **2025**, *13*, 847–885. [CrossRef]
54. Gross, D.; Spieker, H.; Gotlieb, A.; Knoblauch, R.; Elmansori, M. Efficient Milling Quality Prediction with Explainable Machine Learning. *arXiv* **2024**, arXiv:2409.10203. [CrossRef]
55. Jomaa, W.; Mechri, O.; Lévesque, J.; Songmene, V.; Bocher, P.; Gakwaya, A. Finite element simulation and analysis of serrated chip formation during high–speed machining of AA7075–T651 alloy. *J. Manuf. Process.* **2017**, *26*, 446–458. [CrossRef]
56. Ahmadi, S.A.; Davoodi, B.; Niknam, S.A. Numerical modeling and experimental assessment of dynamic behavior of aluminum alloy 7075-T6 in machining process. *Int. J. Adv. Manuf. Technol.* **2024**, *135*, 1131–1149. [CrossRef]
57. Voort, G.F.V. *ASM Handbook: Metallography and Microstructures*; ASM International: Novelty, OH, USA, 2004.
58. Nasr, G.; Soltantarzeh, M.; Davoodi, B.; Hajaliakbari, A. Assessment of tool wear mechanisms in high-pressure jet-assisted turning process of a nickel-based superalloy. *Wear* **2020**, *460–461*, 203454. [CrossRef]
59. Fnides, M.; Amroune, S.; Slamani, M.; Elhadi, A.; Arslane, M.; Jawaid, M. Optimization of Manufacturing Parameters for Minimizing Vibrations and Surface Roughness in Milling Using Box–Behnken Design. *J. Vib. Eng. Technol.* **2025**, *13*, 22. [CrossRef]

60. Slamani, M.; Karabibene, N.; Chatelain, J.-F.; Rimpault, X.; Beauchamp, Y. Statistical analysis of surface roughness during high speed edge trimming of flax-fiber composite based on the Split-Split plot randomized complete block design. *Measurement* **2021**, *184*, 109921. [CrossRef]
61. Dornfeld, D.; Min, S. A Review of Burr Formation in Machining. In *Proceedings of the Burrs—Analysis, Control and Removal*; Aurich, J.C., Dornfeld, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 3–11.

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.