Enhancing Grinding Efficiency in Aluminum Alloys: An Ensemble-Stacking and Single Machine Learning Framework for Predicting Surface Roughness with SHAP-based

The International Journal of Advanced Manufacturing Technology
https://doi.org/10.1007/s00170-026-17594-9

ORIGINAL ARTICLE

Enhancing Grinding Efficiency in Aluminum Alloys: An Ensemble-Stacking and Single Machine Learning Framework for Predicting Surface Roughness with SHAP-based interpretability

Mohsen Dehghanpour Abyaneh[1] · Mohammad Sadegh Javadi[2] · Parviz Narimani[3] · Marzieh Golabchi[4] · Rafaella Sesana[5] · Mohammadjafar Hadad[3]

Received: 22 September 2025 / Accepted: 23 January 2026

Abstract
As mechanical engineering becomes more data-driven, accurate and explainable prediction models are increasingly required. This study applied benchmarks for data-driven strategies and conducted a comprehensive analysis, using an 84-run grinding dataset on aluminum alloy 6061, with the surface roughness as the output variable. The conventional fitting method was the first applied and tuned using a closed-form formula, motivating the use of machine learning algorithms. Gaussian process regression, Artificial neural network, and Extreme gradient boosting were the algorithms that were used in predicting the relationship between input variables and the output. The best single model performance was given by GPR with an accuracy of 97.50%, a MAPE of 2.49%, and an R^2 of 0.99. There were also three models of stacking ensembles that were applied. The stacking ensemble using ANN and XGB as base learners and GPR as the meta-learner offered the best trade-off between its bias and variance and achieved an overall accuracy of 94.54, a MAPE of 5.45, and an R^2 of 0.98. The sensitivity analysis was employed to assess the significance of input parameters. Shapley additive explanations were also used to give attribution to each case, to attribute the impact of individual input features to each prediction. Among the variables, grinding wheel type 89A180K6V111 and the specific removal rate were the most influential. The framework can be adaptable to other datasets and grinding machining scenarios.

Keywords Grinding · Aluminum alloy 6061 · Surface roughness · Stacking ensemble · Shapley additive explanations

1 Introduction

Surface roughness is one of the most important parameters in any type of industrial application, which greatly influences the performance, life cycle, and aesthetic quality of manufactured components [1, 2]. Surface roughness influences not just the functional aspects of the machining process, namely friction, wear resistance, and fatigue life, but also the entire cost-effective performance of the production process in general [3–5]. For instance, aerospace and automobiles manufacturing requires the best surface roughness to ensure the durability and reliability of the products,

✉ Mohsen Dehghanpour Abyaneh
mohsen.dehghanpour@polito.it

✉ Mohammadjafar Hadad
mjhadad@ut.ac.ir

Mohammad Sadegh Javadi
javadims@aut.ac.ir

Parviz Narimani
parviz.narimani@ut.ac.ir

Marzieh Golabchi
marzieh.mohammad@polito.it

Rafaella Sesana
raffaella.sesana@polito.it

[1] Department of Mechanical and Aerospace Engineering (DIMEAS), Politechnico Di Torino, Torino 10129, Italy

[2] Department of Mechanical Engineering, Amirkabir University of Technology, 424, Hafez Ave., P.O. Box, 15875−4413, Tehran, Iran

[3] School of Mechanical Engineering, College of Engineering, University of Tehran, Tehran, Iran

[4] Department of Energy (DENERG), Politechnico Di Torino, Torino 10129, Italy

[5] DIMEAS, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy

Published online: 10 April 2026

especially to withstand high-stress conditions [3, 6]. In addition, precise control of surface roughness has the potential to result in higher performance and reduced requirements for post-processing, which in turn increases the sustainability of manufacturing processes [7]. This means that it is important to understand and control the factors on which the surface roughness depends, including tool geometry, cutting parameters, and properties of the workpiece material, so that high-quality standards of industrial production can be maintained [8]. The level of roughness in the grinding process, particularly for aluminum workpieces, greatly influences the outcome of the final product in terms of quality and performance. Surface grinding parameters allow for an optimum surface finish to be achieved with close control of the grinding parameters, such as wheel speed, feed rate, and depth of cut [9–11]. Aluminum is a soft material and has a propensity to clog the grinding wheel, leading to increased surface roughness and surface defects [12]. To optimize these parameters, it is important to measure the forces during grinding. This is problematic because the forces are small and variable. Proper force measurement will allow us to observe the extent to which the grinding wheel is pressing against the workpiece material that it is cutting. This knowledge can help the grinding machine maintain its condition and have a longer life [13]. Excessive forces may result in more wear and tear on machine components, thus shortening the machine's service life and increasing maintenance costs [14]. This means that force measurement accuracy in measuring and controlling forces is also essential in efficient and effective grinding.

In recent years, new ways of forecasting surface roughness have been developed in the field of machine learning (ML) for grinding processes. A number of ML methods have been used to improve the accuracy and effectiveness of these predictions [15–18]. The use of artificial neural networks (ANNs) in the prediction of surface roughness has been widely used, especially in mechanical processes [19, 20]. For example, research carried out by Balonji et al. [21] applied ANN and adaptive network-based fuzzy inference system (ANFIS) models for forecasting the surface roughness of aluminum alloy 6061 machined blocks. The study showed that the population size, acceleration coefficients, and the selection of membership functions significantly affect the model's prediction accuracy. The hybridization of ANN with genetic algorithms (GA) and particle swarm optimization (PSO) also improved the accuracy of the prediction. Support vector machines (SVMs) have also been looked into as a way to predict surface roughness in mechanical processes. Another ML technique, called extreme gradient boosting (XGBoost), is usually used in the fields of mechanical engineering to guess what will happen with the prediction of the outputs. For example, Shahani et al. [22] created an XGBoost model that uses inputs such as density, wave velocities, and uniaxial compressive strength to predict the static and dynamic Young's modulus of intact sedimentary rocks. The model performed better than ANNs and conventional regression techniques, and showed high accuracy. Similarly, Mustapha et al. [23] carried out a study with the help of XGBoost. This algorithm was applied to predict the mechanical and physical properties of pervious concrete, such as compressive strength, tensile strength, density, and porosity. The model reached high predictive accuracy, which demonstrates robustness in modeling complex material behavior. Ou et al. [24] used an ensemble of XGBoost and discrete element modeling to predict autogenous-grinding mill throughput. The model's results were in good agreement with the validation dataset. XGBoost achieved a good predictive performance on the test dataset. Gao et al. [25] proposed a model of material removal using ML and acoustic sensing for robotic-belt grinding of Inconel 718. A new k-fold Extreme Gradient Boosting (k-fold-XGBoost) method was used for the material-removal (MR) model after the training and optimization process. The test results indicate that the model-predicted values and measured values are in close agreement. According to the model, the mean absolute percentage error (MAPE) for material removal was 4.37%. Samavatian et al. [26] used three different algorithms, Gaussian process regression (GPR), gradient boosting (GB), and K-nearest neighbors (KNN) regression for determining surface roughness in steel grinding processes. The GPR model performed better than the other models. According to the results, GPR achieved the highest predictive accuracy with an R^2 value of 0.96, compared to 0.93 for GB and 0.90 for KNN.

Simultaneously with the practice of grinding, ML has quickly been embraced in the prediction of surface roughness in other highly precise and ultra-precision machining processes [27]. In ultra-precision milling, Shang et al. [28] constructed an extreme learning machine (ELM) model using sensor-data fusion to predict surface roughness in ultra-precision milling of optical components, and they showed that data-driven models can outperform classical regression in the case of cutting parameters and vibration measurements fused. Many process parameters, such as spindle current, vibration, and acoustic emission, were used in this study to predict surface roughness. The results indicated that the suggested approach had high prediction accuracy. This indicated the effectiveness of SVMs for predicting surface roughness. Bai et al. [29] suggested a hybrid physics-data-driven model that integrates mechanistic descriptors with ML surrogates for ultra-precision machining, with high predictivity and at the same time physical interpretability of the underlying processes. Ruan et al. [30] presented NASPrecision, a neural-architecture-search-based,

multi-stage learning scheme that automatically identifies the best DL architectures in predicting surface roughness in multiple ultra-precision machining datasets at the framework level, finding much smaller MAPE and RMSE than multi-model baselines.

Similar deep learning (DL) models have been suggested for online monitoring.The application of DL models, both convolutional neural networks (CNN) and long short-term memory (LSTM) networks, has shown potential in predicting surface roughness [31, 32]. A study by Wang et al. [33] focused on predicting surface roughness during large-shaft grinding. It proposed an attentional CNN-LSTM architecture. The model combined several process signals and a self-attention mechanism to automatically assign weights to the input features. The final experimental results showed that the suggested technique delivered better prediction accuracy than traditional methods. Another study by Guo et al. [34] used an LSTM network to predict ground surface roughness using grinding force and acoustic emission signals. The findings indicated that the LSTM model performed well, which means that DL models are promising in this field. Shehzad et al. [35] came up with an integrated hybrid CNN-LSTM system framework to predict the quality of the surfaces in ultra-precision fly-cutting of copper workpieces, whereby tooltip acceleration signals are used as inputs and obtained better results in comparison to conventional CNN or LSTM models. The same group applied a study later that proposed a DL-aided online monitoring system of surface roughness in ultra-precision fly-cutting, which validated the practicability of complete information-driven surveillance in industrial-grade precision machining. However, DL models require a large-scale dataset to attain optimal performance. Therefore, the applications of DL models would be limited in scenarios like experimental studies where there is only a limited dataset available.

In addition to fly-cutting, Sizemore et al. [36] used ML to categorize the quality of surfaces during ultra-precision diamond machining of single-crystal germanium and demonstrated that supervised classifiers can be effectively used to distinguish between acceptable and defective surfaces and, therefore, to inform parameter selection in optical manufacturing. End-milling has also been modeled using soft-computing and ML models, such as the example of Balonji et al. [37], who applied multiple soft-computing methods to predict Ra in CNC end-milling of AA6061, and where optimized models have high R^2 values and low RMSE.

Electric discharge machining (EDM) and wire EDM (WEDM) also have a high degree of non-linear behaviour, and a number of studies have shifted to ML-based prediction of surface integrity. Abbas et al. [38] predicted the surface roughness of die-sinking EDM of composites based on aluminum matrices with Decision Trees (DT), Random Forests (RF), and XGBoost, with better prediction performance than classical response-surface models. Gurgenc and Altay [39] forecasted the surface roughness of WEDM in AZ91D magnesium alloy with the help of Multilayer Perceptron (MLP), ensemble neural networks, and evolving product-unit neural network, and have concluded that ensemble and evolutionary neural networks are better predictors than traditional MLPs, and they require significantly less experimental work.

Over the last few years, there have been major breakthroughs in ML that enhance the accuracy of models and their predictive potential. One such approach is the stacking ensemble method. Baraheni et al. [40] applied a stacked ensemble to forecast delamination and thrust in 72 carbon-fiber-reinforced polymers (CFRP) drilling runs. The superiority of the ensemble over the individual models was evidenced by a significant improvement in the MAE and RMSE values, up to 97% and 124% in the case of delamination, and 205% and 154% in the case of thrust force, compared to the best base learner. Natarajan et al. [41] proposed a two-stage pipelined architecture with discretization of the output variable in which a regressor is driven by a classifier. The resulting meta-model was highly accurate ($R^2 > 90\%$ for the XGBoost regressor against all materials) and greatly outperformed single models in predicting Ra. Jones et al. [42] combine the load cell, accelerometer, and acoustic-emission signals into a three-level architecture and stacked RF, ANN, and XGBoost to predict CNC milling tool flank wear. The ensemble obtained an R^2 of 0.95 and an RMSE of 58.23 μm, which was superior to any single model and facilitated prognostics and health management in real-time monitoring of machining.

The sensitive input variables available in previous studies also often fail in documenting their impacts on individual forecasts, determining the relative significance of input parameters under different circumstances, and explaining why specific inputs are relevant to variations in surface roughness. Interpretable ML models can provide vital information on the intricate nonlinear dynamics of input parameters throughout the grinding process. To this end, the paper examines the behavior of surface roughness in the grinding process, through Shapley Additive Explanations (SHAP). SHAP can be applied to interpret various ML models and assign importance scores to each input variable for a given prediction. The experimental results have not, however, been used to derive an interpretable ML model in previous studies, i.e., a model that identifies the contribution of each input to a particular prediction. The research by Huang et al. [43] employed SHAP to explain surface roughness model during the milling of thermally modified timber. Several ML models, including RF, SVM, GPR, and MLP were implemented in this investigation. The model that performed best

was the GPR model with R^2 =0.96 and RMSE = 1.59 µm. SHAP found spindle speed and feed rate to be the most influential feautures. Timilsina et al. [44] applied ML models, e.g., Extra Trees, RF, XGBoost, and SVR, to predict the higher heating value (HHV) of waste-derived fuels. They explained local and global feature importance using SHAP. The models had a high level of accuracy, with R^2 between 0.83 and 0.98, RMSE ranging from 0.79 to 2.25, and MAPE between 0.92% and 6.01%. SHAP analysis revealed that carbon and hydrogen were the most effective features, with mean SHAP values of 2.17 and 0.37, respectively. Pašić et al. [45] applied ANN and SHAP to analyze the turning of 90MnCrV7 steel. SHAP indicated that the insert radius was the most important factor among the output parameters. In their work, they found the optimal cutting conditions to reduce surface roughness and forces. Although SHAP has been used in other applications [46–49], no study has thus far applied SHAP in the analysis of ML prediction of surface roughness in the grinding process. The current study addresses this gap directly.

Our earlier research on grinding can be seen in the context of this larger picture of ML-assisted ultra-precision machining. To regulate the surface roughness in sustainable grinding of the St37 steel, we initially employed ANNs along with optimization to control the surface roughness at varying dressing and cooling conditions [50]. We then applied this approach to nickel-based superalloy Inconel 738, and introduced more dressing-tool geometries and deeper ANN architectures to predict surface roughness, and to UNS S34700 stainless steel, where ANN, GPR, and SVR were incorporated together with genetic algorithms to predict both surface roughness and grinding forces, and explicitly model the three types of wheels and four used coolants [51, 52]. Relative to these previous grinding-concentrated investigations and even, as compared to the above attempts in turning, milling, fly-cutting, and EDM, the current research appears to be the first to utilize a stacking-ensemble framework with SHAP-based explainability to the prediction of surface roughness in sustainable grinding of AA6061. It thus offers an interface between state-of-the-art ensemble learning, explainable ML, and the larger movement of ML-assisted ultra-precision machining. Table 1 presents a comprehensive framework of current research utilising ML techniques in ultraprecision and high-precision machining processes.

In the current study, single ML methods, stacking ensemble algorithms, sensitivity analysis, and SHAP interpretation are utilized to understand how the specific removal rate, coolant type, and grinding wheel type significantly influence surface roughness. This paper demonstrates that research on the potential of hybrid ML models that incorporate the combination of various techniques can result in a considerable increase in prediction accuracy.

2 Methodology

2.1 Experimental setup

This study extends previous experimental research [53, 54], concentrating on AA6061 aluminum alloy with the M7135A NANTONG SHUANGZANG-NC surface-grinding machine in down-grinding (plunge) mode. Some tests were performed on the workpiece to determine the best lubricant range and compressed air flow range by synthetic and vegetable oil. Abrasive types, grain sizes, and level of hardness varied among the three types of grinding wheels used in this study. This cooling process was done through four combinations of coolants, namely dry compressed air, MQL using two vegetable oils of different viscosities, and water-miscible fluid, all utilized at seven depths of cut. The specific removal rate is proportional to the work speed and the depth of cut. As mentioned before, seven specific removal rates were considered during the test process, namely 5, 15, 25, 35, 45, 55, and 65 mm3/(mm.min). Four types of coolant were used: dry compressed air, MQL1 and MQL2 (different based on oil types), and Fluid type. In addition, three grinding wheels served as input parameters. For grinding wheel types, aluminum oxide 89 A is suitable for high-strength applications such as alloy steels, whereas 88 A is used for general-purpose grinding. With grinding wheel 89A180K6V111, a harder grade (K) and a superfine grit were used. Silicon carbide (C) was used in another grinding wheel type, which is sharper and more brittle than aluminum oxide. The output measured parameter was surface roughness. All other grinding conditions were kept constant, as detailed in Table 2.

Since the AA6061 material is adhesive-prone and prone to wheel loading, special attention was paid to the control of the state of the grinding wheel during the experiments. In the initial experimental campaign, from which the current data set will be taken, all tests were carried out with TYROLIT wheels 89A180K6V111, 88A80L6AV217, and C120I6AV1850, all dressed with a single-point diamond dresser to a total depth of 40 mm at the dressing speed of 150 mm/min (Table 2), as stated in the earlier study [43]. In every grinding condition (each of the combination of wheel type, coolant, and a given removal rate), the grinding was conducted in the plunge mode and the surface roughness and the forces of grinding were measured after the tenth pass and consequently, all the data points in the current work are connected with the same total accumulation of the active wheel surface [43]. All the other machine parameters (wheel speed, table speed, nozzle position, and coolant delivery parameters) were held constant such that the wheel topography developed in a similar manner throughout the test matrix [53, 54]. The previously conducted experiment

Table 1 A general framework of existing research employing ML techniques in machining processes

Ref.	Study Focus	Material	Machining Process	Key ML Technologies	Reported Accuracy Metrics
[26]	Ra prediction with analytically guided	Various steel types	Surface grinding	Physics-informed features + ML models	Higher R^2 and lower RMSE than purely empirical models
[28]	Ra prediction with sensor-data fusion	Optical components	Ultra-precision milling	ELM with data fusion	Lower error than classical regression; improved prediction
[29]	Ra prediction with hybrid physics–data model	Titanium alloy workpieces	Ultra-precision machining	Physics-informed ResNet-based ML surrogate	High predictive accuracy and preserved physical interpretability
[30]	Ra prediction with the NASPrecision framework	Multiple ultra-precision machining datasets	Ultra-precision machining	NAS-driven multi-stage deep networks	Significant reductions in MAPE and RMSE vs. single-stage baselines
[35]	Online Ra monitoring	Copper workpieces	Ultra-precision fly-cutting	Hybrid CNN–LSTM using tooltip acceleration time series	Lower prediction error and higher accuracy; robust online monitoring
[36]	Surface quality classification	Single-crystal germanium	Ultra-precision diamond machining	Supervised ML classifiers (ANN, SVM, KNN, DT, RF)	High classification accuracy; reliable separation of acceptable vs. defective surfaces
[37]	Ra prediction	Al6061 (aluminum alloy)	CNC end milling	ANN-based soft-computing techniques	High R^2 and low RMSE for optimized models
[38]	Ra prediction	Al/SiC/Gr hybrid composite	Die-sinking EDM	DT, RF, XGBoost	Improved prediction vs. response-surface models; better fit to non-linear behaviour
[39]	Surface roughness prediction	AZ91D magnesium alloy	Wire EDM	MLP, ensemble neural networks, evolving product-unit NN	Ensemble/evolutionary NNs outperform standard MLP; reduced experimental effort
[50]	Ra control under multiple dressing and cooling conditions	St37 mild steel	Surface grinding (sustainable grinding)	ANN with optimization (various architectures)	R^2, RMSE, MAPE, Accuracy, good control under four cooling strategies
[51]	New parameters for advanced surface roughness prediction	Inconel 738	Surface grinding	Deep MLP / ANN with extended parameter set	MSE, R^2, RMSE, MAE; improved prediction with additional process descriptors
[52]	Joint prediction of Ra and grinding forces with coolant effects	UNS S34700 stainless steel	Surface grinding	ANN, GPR, and SVR combined with Genetic Algorithms	R^2, RMSE, MAPE, and Accuracy
This work	High-fidelity Rz prediction with interpretability	AA6061 aluminum alloy	Surface grinding (sustainable, multi-coolant, multi-wheel)	Stacking ensembles, ML models with SHAP-based interpretability	R^2, RMSE, MAPE, Accuracy; first stacking-ensemble SHAP study in sustainable grinding of AA6061

indicated that AA6061 is prone to redeposited coatings and loading chips on the wheel, particularly in synthetic ester MQL, unlike the SiC wheel (C120I6AV1850), which displays self-sharpening properties and less redeposition on the work surface [43]. The effects of the grinding tool condition are controlled and systematic by maintaining a consistent dressing procedure, constant number of passes per condition and restricting the experiments to the operating window in which it is possible to observe constant forces and quality of surface and so the changes in Rz discussed in this paper are dominated by the effects of the coolant type, wheel specification, and removal rate, rather than random alterations in the wear of the wheel or wheel loading.

The choice of the variable parameters in Table 2 was not random but based on our previous studies on grinding experimental design and improved in pre-experiments. The

Table 2 Grinding properties parameters includes fixed and variable parameters

	Grinding Condition	Types	More details
Fix Parameters	Grinding mode	Down surface grinding (plunge)	-
	Grinding machine	M7135A-NANTONG SHUANGZANG	-
	Wheel speed (V_C)	30 m/s	-
	Work Speed (V_{fi})	1500 mm/min	-
	Workpiece material	Aluminum 6061 with 170 HV	-
	Dresser	Single point diamond dresser	-
	Total depth of dressing (a_d)	40 µm	-
	Dressing speed (V_d)	150 mm/min	-
Variable Parameters	Depth of cut (a_e)	5, 15, 25, 35, 45, 55, and 65 µm	-
	Coolant	Dry MQL1 MQL2 Fluid	- synthetic ester oil, ASTM D-445 = 23.96, Q = 100 ml/h; P = 4 bar vegetable oil, ASTM D-445 = 38.6, Q = 100 ml/h; P = 4 bar Water miscible (based on mineral oil in a 5% concentration)
	Grinding wheel type	89A180K6V111, 88A80L6AV217, C120I6AV1850	Manufactured by TYROLIT Co., with d_s=400 mm

original experiment on MQL grinding of AA6061 and UNS S34700 represented a series of screening experiments that needed to be performed to find suitable ranges of depth of cut, specific removal rate and also to avoid grinding burn or chatter or unstable forces; the final set of seven removal rates (5–65 mm depth of cut, 7.5–97.5 (mm3/(mm*min)) that permit the generation of measurable and systematic variations in Rz and forces without damaging surface integrity was thus discovered. Similarly, the MQL conditions (synthetic ester and vegetable oils at 100 ml/h and at 4 bar, nozzle distance 100 mm) were also set following preliminary experiments with each of the workpiece materials to establish the most suitable lubricant type, oil flow, and air pressure in the wider MQL program of Hadad and co-workers of temperature, energy partition, and grindability [52, 53]. The current work is simply a modification of the depth of cut (i.e. a particular removal rate) of the cut, the type of coolant, and the type of wheel within this pre-validated window, with all other parameters held constant, meaning that the ML analysis concentrates on the physically most significant variables revealed in the prior experimental campaign.

The process measurements include surface roughness, which was reported individually for each test. Workpiece roughness was measured using a DIAVITE AG mobile roughness measurement device with a cut-off length of 0.8 mm, in accordance with DIN EN ISO 3274:1998 standards. At the end of each test, surface roughness (Rz) was measured at five different points along the grinding direction. The grinding force components were recorded using a four-component dynamometer (Kistler type 9272) positioned under the workpiece clamping device (Fig. 1).

2.2 Dataset evaluation

There are several parameters involved in a typical grinding process; however, considering all the influential parameters requires a complex testing procedure, which is costly and may yield results that are not universally applicable. Therefore, due to the direct impact of key parameters on grinding quality and machine longevity, the following parameters are considered: specific removal rate, coolant type, and grinding wheel type. In this article, these parameters are referred to as removal rate, coolant, and wheel type for simplicity.

The incorporated dataset includes 84 sets of data, which are shown in Table 3, representing a subset of the full dataset. Table 3 is divided into input and output parameters, commonly referred as predictors and response. In Table 3, the grinding wheel types are 89A180K6V111, 88A80L6AV217, and C120I6AV1850, which are denoted A, B, and C, respectively.

2.3 ANOVA

Analysis of Variance (ANOVA) is a statistical technique employed in analysing the variations between group means in a sample. It is especially applicable in mechanical engineering to determine the important variables and interactions in experiment design, including optimal material and processes, manufacturing, and performance characteristics. ANOVA tends to be employed along with robust design in order to identify the parameters of a device that influence its performance to the greatest extent during its development process [55]. The ANOVA procedure usually starts with a complex model that takes into account all main effects and interactions. To make the regression model easier to understand without losing too much quality, factors with the

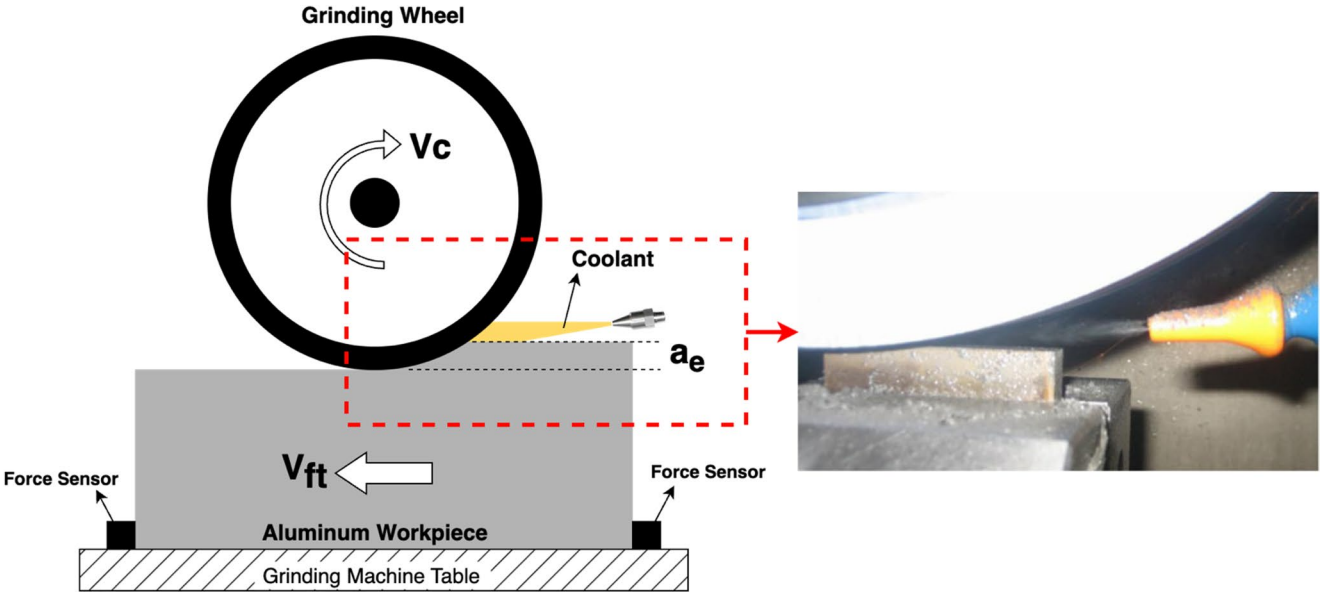

Fig. 1 Illustration of the testing process

Table 3 Grinding properties parameters dataset at a glance

Input Parameters			Output Parameter
Specific removal rate (mm^3/mm.min)	Coolant type	Grinding wheel type	Rz (µm)
52.5	MQL2	A	2.11
7.5	Fluid	A	1.7
97.5	Fluid	A	2.15
22.5	MQL1	B	1.1
82.5	MQL2	B	1.36
67.5	Fluid	B	0.93
97.5	Dry	C	1.82
52.5	MQL1	C	1.12
82.5	MQL1	C	1.34
97.5	MQL1	C	1.45
7.5	Fluid	C	0.87
22.5	Fluid	C	0.93

highest p-values are taken out one by one. This makes sure that the parameters with the lowest p-values, which have the most effect, remain [56].

In the present research, where more than two factors are involved, N-way ANOVA is employed with n representing the independent variables in the analysis. N-way ANOVA provides an opportunity to examin the interaction of multiple factors in detail and their overall effect on the dependent variable, which gives a better understanding of the underlying dynamics. The inclusion of N-way ANOVA was intended to highlight how advanced statistical approaches can be used to evaluate the significance of multiple factors simultaneously. Compared to traditional response surface methodologies (RSM), N-way ANOVA offers several advantages:

- Simultaneous evaluation of multiple factors and their interactions.
- Clear statistical validation by providing p-values.
- Robustness with limited datasets.
- Transparency and interpretability.

Therefore, this technique gives ANOVA results that are straightforward to interpret [57–59].

2.4 Particle swarm optimization

Particle Swarm Optimization (PSO) is a population-based stochastic optimization algorithm inspired by the intelligent collective behaviour of animals, such as flocks of birds. It was first introduced in 1995 [60]. The main design concept of the PSO algorithm is closely related to two areas of research: evolutionary algorithms and artificial life. Like evolutionary algorithms, PSO uses a swarm model that allows it to simultaneously search a large region in the solution space for the optimized objective function. Artificial life, in turn, explores systems that exhibit life-like characteristics.

In the field of grinding process models and computer-aided manufacturing, PSO provides a basis for optimizing grinding parameters. Numerous variables affect the economics of machining operations, including machine tool capacity, required workpiece geometry, and cutting conditions such as speed, feed, and depth of cut [61]. PSO offers a robust solution for finding the optimal relationship between these variables, particularly in grinding applications. For instance, it can calculate the minimum production cost depending on the machining conditions [61] and reach the

highest material removal rate (MMR) with the maintenance of surface finish and limitation of surface damage [62].

2.5 Formula evaluation

Given the numerous process variables involved in grinding and the sensitivity of changes to the final performance of parts, optimizing every variable is both difficult and complex. Fortunately, some of these variables are determined by operators, and some are more critical than others. This part of the study focuses on the effects of removal rate, coolant type, and grinding wheel types, aiming to find the analytical formulation between variables related to surface roughness and axial forces.

Two main concerns arise from this overview. First, the coolant type and grinding wheel are qualitative parameters that need to be converted into appropriate quantitative variables for any analytical formula. This conversion is explained in detail in the Results and Discussion section. Second, the structure of the evaluation formula is crucial. While there are countless possible relationships between input parameters and their coefficients, for this study, the relationship between single and dual variables with appropriate coefficient combinations is deemed sufficient.

Another important point is that the combination of variables and associated coefficients, treated as unknown parameters to be solved by PSO, should be less than the number of data points available in the dataset. This is because PSO uses the dataset as a reference to solve the optimization problem, and reaching a unique solution requires satisfying this condition. Therefore, the formulas illustrated in Eq. 1 are considered evaluation relationships between input parameters and surface roughness output.

$$\begin{gathered} R_z = a_1 x_r + a_2 x_c + a_3 x_w + x_r^{a_4} + x_c^{a_5} + \\ x_w^{a_6} + \frac{a_7}{x_r} + \frac{a_8}{x_c} + \frac{a_9}{x_w} + a_{10} x_r x_c + a_{11} x_c x_w + a_{12} x_r x_w + \\ a_{13} \frac{x_r}{x_c} + a_{14} \frac{x_c}{x_w} + a_{15} \frac{x_r}{x_w} + \frac{a_{16}}{x_r x_c} + \frac{a_{17}}{x_x x_w} + \frac{a_{18}}{x_r x_w} \end{gathered} \quad (1)$$

In the above equation, (x_r) represents the removal rate, (x_c) denotes the coolant type (defined by a specific number), and (x_w) indicates the grinding wheel type (also defined by a specific number for each wheel type). The coefficients a_i represent the weights assigned to each term in the proposed surface roughness formulas. These coefficients are individually optimized using PSO.

2.6 Machine learning

2.6.1 Gaussian process regression

Gaussian process regression (GPR) follows a distribution function defined by its mean and covariance functions [63]. The GPR model is defined over a given set of vector-valued data to predict the output based on a distribution function estimated from the data with a specified covariance function [64]. For this reason, according to Eq. 2, f(x) is defined as a weighted sum of basis functions:

$$y(x) = f(x) = \sum_{i=1}^{N} w_i \varphi_i(x) + \sigma_f \epsilon = W^T \varphi(x) + \sigma_f \epsilon \quad (2)$$

Where ϵ represents a white noise with correlation defined by σ_f across noise models, W denotes the weight matrix associated with the output, and $\varphi(x)$ gathers the values of the N basis functions evaluated at x.

The covariance function, which is known as the kernel function, is also a method for mapping input data (linear or non-linear) into a feature space (Eq. 3). The function values corresponding to any two inputs are shown below.

$$cov(f(x), f(x')) = \varphi(x)^T \sum W \varphi(x') = k(x, x') \quad (3)$$

In the end, the likelihood for a given function (f) is defined, and solving it provides the optimal set of weight values while considering system noise and uncertainties. This model also offers a robust solution.

This model is a popular and powerful ML example, enhanced by a Bayesian approach to provide uncertainty evaluation of the predicted values. In mechanical engineering, GPR has demonstrated robust performance with small datasets when predicting surface roughness. It provides a model based on input parameters such as depth of cut, wheel velocity, feed, and the velocity of the workpiece [64]. Furthermore, measuring and analyzing cutting force has a significant impact on the metal turning process. GPR is also an effective method that can ensure high predictive accuracy for cutting force prediction during the turning of AISI 4340 steel [65]. Figure 2 illustrates the GPR methodology used in this study. Some well-known kernel functions are shown in Table 4. The Radial Basis Function (RBF) kernel is a general category, with the Squared Exponential (SE) kernel being its most common variant. There are also several kernel functions used in GPR across various applications [66, 67].

2.6.2 Artificial neural network

Artificial Neural Networks (ANNs) have been successfully applied to a variety of application problems, including classification and function approximation. They are particularly useful as function approximators because they do not require prior knowledge of the input data distribution and have been shown to be universal approximators [68].

The implemented ANN structure is based on a Multi-Layer Perceptron (MLP), a feed-forward neural network that maps sets of input data to appropriate outputs. This structure consists of an input layer, a single-node output layer, and a single hidden layer consisting of a varying number of nodes

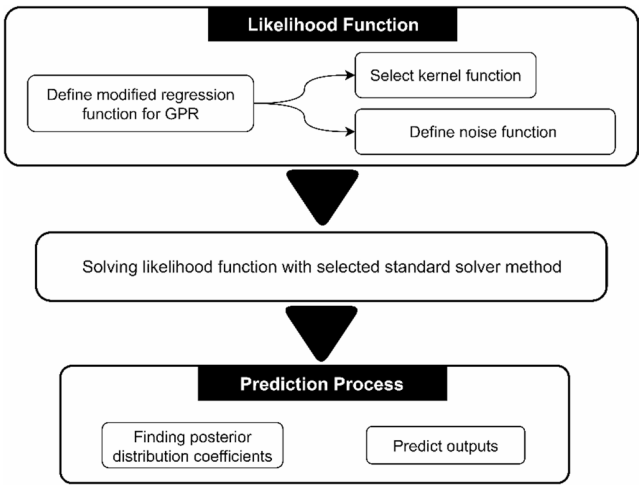

Fig. 2 Schematic illustration of GPR modelling

(Fig. 3) [69]. In ANN, activation functions are crucial as they aid in learning and understanding non-linear and complex, non-linear relationships between inputs and outputs. Although more than ten activation functions are discussed in the literature [70], the ReLU activation function has demonstrated adequate performance for the current study.

ANNs, like other ML methods, provide excellent performance in predicting complex relationships between related variables. The scalability and flexibility of ANNs allow them to be adapted to a wide variety of problems, especially in mechanical engineering. For example, ANNs can be adapted to investigate batch grinding conditions using a kinetic model to predict the grindability of calcite [15]. Additionally, they can model energy consumption, which varies with grinding wheel wear over its lifecycle by implementing time-series ANNs to support a more efficient grinding process [71].

Table 4 Some frequently used kernel function for GPR

Kernel Function Name	Standard Formula	Notes
Constant Kernel	$K\left(x_i, x_j\right) = \sigma_f^2$	Adds a constant offset; often combined with other kernels.
White Noise Kernel	$K\left(x_i, x_j\right) = \sigma_n^2 \delta_{ij}$	Useful for handling measurement errors, sensor noise, or aleatoric uncertainty in regression problems
Squared Exponentia (RBF)	$K\left(x_i, x_j\right) = \sigma_f^2 exp\left[\frac{-1}{2}\frac{\left(x_i - x_j\right)^T\left(x_i - x_j\right)}{\sigma_l^2}\right]$	Ideal for smooth functions and homogeneous data
Rational Quadratic	$K\left(x_i, x_j\right) = \sigma_f^2\left[1 + \frac{\left\|\left(x_i - x_j\right)^T\left(x_i - x_j\right)\right\|}{2\alpha\,\sigma_l^2}\right]^{\alpha}$	For modeling complex patterns that vary across input space.

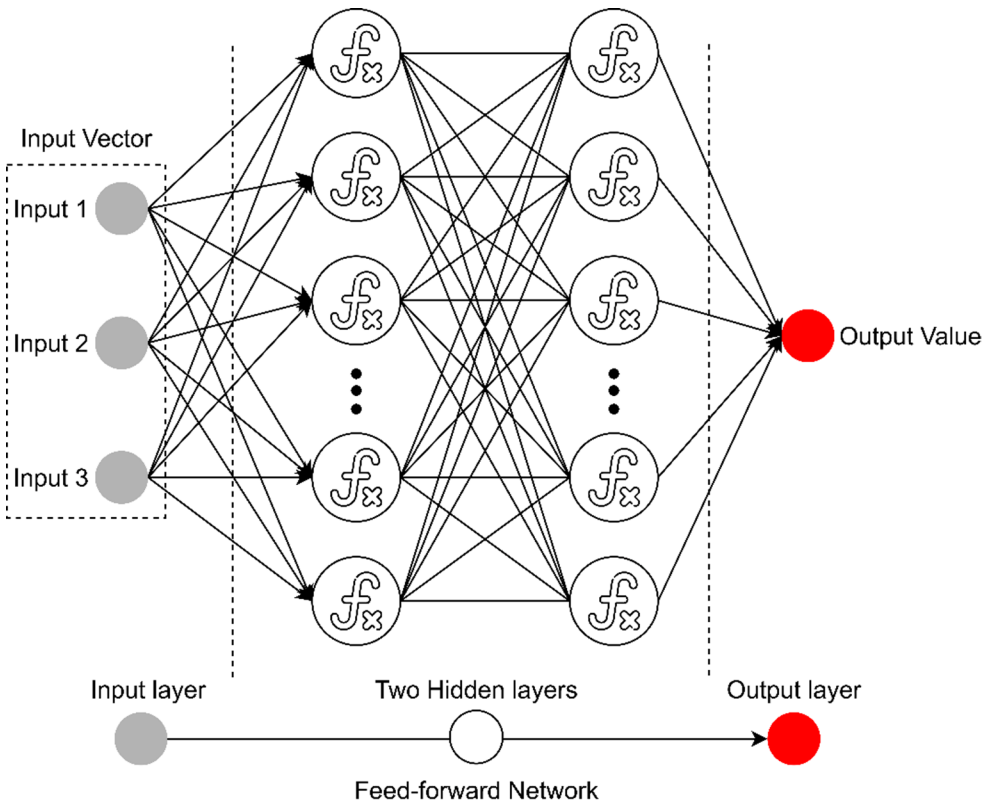

Fig. 3 Schematic illustration of the ANN model

2.6.3 Extreme gradient boosting

Extreme Gradient Boosting (XGBoost) is a powerful approach for merging decision trees that is often used in classification and regression applications. In RF, trees develop separately from each other, but in XGBoost, trees are formed one after the other, with each tree fixing the mistakes of the one before it. In this study, the XGBoost package from Python was used. Unlike GPR, which models data using probabilistic distributions defined by mean and covariance functions, XGBoost focuses on maximizing a regularized objective function to enhance prediction accuracy while minimizing overfitting [42]. In XGBoost, each tree corrects the mistakes of the preceding tree using a loss function (Eq. 4):

$$\boldsymbol{y_i} = \sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \boldsymbol{f_t}\left(\boldsymbol{x_i}\right) \tag{4}$$

where x_i is just the feature vector for the specimen i, y_i is the true label or target value for specimen i, and f_t is the prediction from the t-th tree.

One of the most important hyperparameters for decision trees in an XGBoost algorithm is the max depth per tree. The deeper the trees, the more complex the patterns can be. It can lead to overfitting if the max depth is very high. The maximum depth is defined by the following function (Eq. 5):

$$\boldsymbol{d_{max}} = \boldsymbol{max}\left(\boldsymbol{depth}\left(\boldsymbol{f_t}\right)\right) \tag{5}$$

In the above, d_{max} is the maximum depth of any tree f_t in the model. The parameter γ is the learning rate controlling each interaction's learning rate. In order to converge on more trees, a smaller learning rate is typically required. This makes the training process longer, resulting in better generalization. Equation 6 defines the updated prediction y_i at iteration t, incorporating the learning rate γ.

$$\boldsymbol{y_i^{(t)}} = \boldsymbol{y_i^{(t-1)}}\boldsymbol{\gamma}\ \left(\boldsymbol{f_t}\left(\boldsymbol{x_i}\right)\right) \tag{6}$$

Here, $y_i^{(t)}$ is the prediction after incorporating the *t*-th tree into the ensemble. The subsample hyperparameter controls how much training data is used to build each tree, which is important to keep from overfitting. By adding randomness, lower subsample values make models more stable and lower the risk of overfitting. Subsampling is a kind of stochastic gradient boosting, and its effect can be expressed as follows (Eq. 7):

$$\boldsymbol{N_{tree}} = \boldsymbol{s}\left(\boldsymbol{N_{total}}\right) \tag{7}$$

Here, N_{tree} is the number of samples for training each tree, and N_{total} is the overall number of training samples in the datasets. The parameter s is the subsample fraction, which is a value between 0 and 1, indicating the proportion of data used per tree. The fraction of features is controlled by the column sample hyperparameter. To prevent overfitting, the column sample hyperparameter was set to 1.0, meaning each tree has access to all features, which represents the number of features available to each tree. In a dataset with M features, each tree uses the following number of features (Eq. 8):

$$\boldsymbol{M_{tree}} = \boldsymbol{c}\left(\boldsymbol{M_{total}}\right) \tag{8}$$

Here, M_{tree} refers to the number of features selected to train each tree, and M_{total} is the overall number of features from the dataset. The parameter c is the column sampling rate, which is a value between 0 and 1, indicating the proportion of features used per tree Because its strong performance and flexibility, this ML method is one of the most widely used approaches for predictive tasks in mechanical engineering, such as tool wear prediction [72], surface quality assessment [73], and machining parameter optimization [73]. Figure 4 illustrates the architecture of the XGBoost model employed in this study.

2.6.4 Stacking ensemble

Stacking is a well-established ensemble learning technique used in both regression and classification tasks [74]. First introduced by Wolpert (1992) under the term "Stacked Generalization", the method enhances predictive accuracy by combining the outputs of multiple models, then, Breiman (1996) proposed internal K-fold cross-validation for the same form of stacking [75, 76]. This model combines the predictions made by several basic models to generate a prediction that is more reliable and accurate. Stacking ensembles can be applied to challenges involving regression and classification. Several base models (ML models) are trained on the same dataset in a stacking ensemble. A higher-level model, often known as a meta-model, uses the base models' predictions as inputs or training data. By utilizing the advantages of the base models, the meta-model gains the ability to make a final prediction [77]. Figure 5 presents the flowchart of the stacking ensemble architecture used generally. The stacking ensemble architecture in this research is unique as it has two base-learner ML models and one meta-learner ML model. The two base learners are trained using the dataset and make Rz predictions on a test set. The meta-learner is trained on data consisting of Rz predictions made by the base-learners. The meta-model does not utilize the original 28 statistical

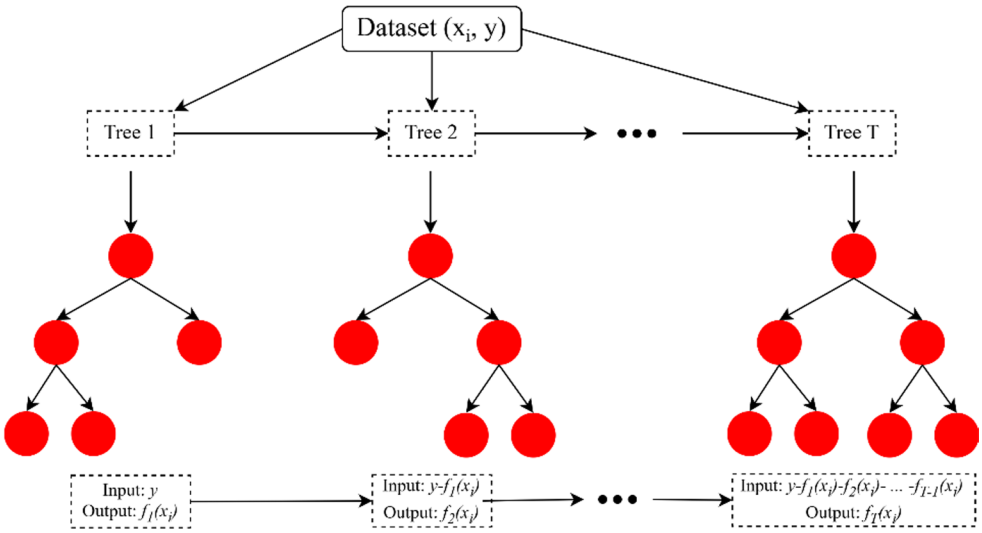

Fig. 4 Schematic illustration of the XGBoost model

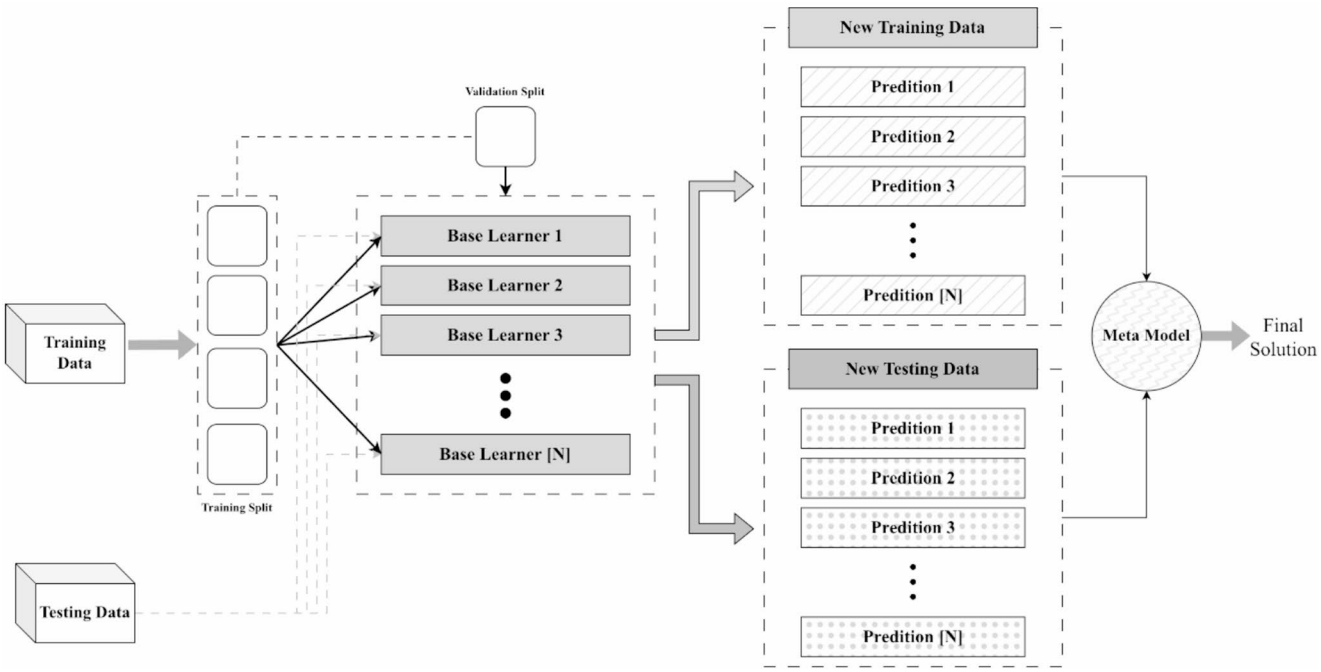

Fig. 5 Flowchart of a stacking ensemble process

features, but utilizes the base model predictions as features. In order to provide a fused forecast that is more accurate, the meta-model is validated on a test set. Decision-level fusion is possible by using the stacking ensemble ML method. Overall, stacking offers a flexible and powerful strategy for improving both the reliability and precision of ML models, particularly in complex problem domains [78, 79].

2.7 Accuracy metrics

In the regression model, especially a nonlinear structure, which is a complex relation between independent values and target value, several metrics are available; some studies discuss up to 29 different evaluation metrics [80]. For just regression, however, using all of them to select the best model is often unnecessary. In the real world, comparing metrics, which reflect model stability and accuracy, gives complete insight about performance, which introduces three main metrics that the current study used. The root mean squared error (RMSE) is a widely used metric for evaluating models, particularly for normal (Gaussian) errors [81]. The RMSE is the square root of the mean squared error (MSE). Taking the root does not affect the relative ranks of models, but it yields a metric with the same units as the input data, which conveniently represents the typical or "standard" error for normally distributed errors (Eq. 9).

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{\boldsymbol{n}}\sum_{1}^{\boldsymbol{n}}(\boldsymbol{Y_{pre}} - \boldsymbol{Y_{act}})^2} \tag{9}$$

Mean Absolute Percentage Error (MAPE) is another metric used to measure the extent of absolute error in percentage terms. This metric is frequently used as a loss function in various studies (Eq. 10), representing the average percentage error in the prediction process [82].

$$\boldsymbol{MAPE} = \frac{1}{\boldsymbol{n}}\sum_{1}^{\boldsymbol{n}}\frac{|\boldsymbol{Y_{pre}} - \boldsymbol{Y_{act}}|}{\boldsymbol{Y_{act}}} \times 100 \tag{10}$$

The Coefficient of Determination, often denoted as (R^2), becomes quite complex when applied to nonlinear regression models or other nonlinear approaches, where the value does not always fall between 0 and 1; it can even be negative. This is a major drawback to the reliability of R^2 in nonlinear regression, which has been raised by various scholars. Kvålseth [83] investigated eight different formulas for R^2 found in the literature, highlighting their differences, the confusion they cause. They also discussed common mistakes in their application. However, the standard formula (Eq. 11) serves as a measure of "goodness of fit" and, when used alongside other metrics, provides a good understanding of the implemented model's stability and accuracy [84].

$$\boldsymbol{R^2} = 1 - \frac{\sum_{1}^{\boldsymbol{n}}(\boldsymbol{Y_{act}} - \boldsymbol{Y_{pre}})^2}{\sum_{1}^{\boldsymbol{n}}(\boldsymbol{Y_{act}} - \overline{\boldsymbol{Y}}_{\boldsymbol{act}})^2} \tag{11}$$

In the above Eqs. 8, 9, and 10, Y_{act} is the actual value, Y_{pre} is the predicted value, $\overline{Y}_{act}$ is the mean value of the actual output, and $\overline{Y}_{pre}$ is the mean value of the predicted output.

2.8 Sensitive analysis

Sensitivity analysis is important in advanced machining processes like grinding to have knowledge of the positive or negative influence of input parameters on output responses. This is a useful tool that is applied in determining how variation of levels of the input parameters impacts the output, such as the surface roughness. Besides, the figure enables comparison of the experimental findings with ML techniques that point out predictive accuracies of various ML models. This type of approach enables better estimation of the process behavior and leads to improved strategies for optimizing it through predictive analytics. It enables engineers and researchers to make more data-driven decisions that would make the quality of products more desirable and create stronger manufacturing systems.

2.9 Shapley additive explanations

Shapley Additive Explanations (SHAP) is a game-theoretic approach that represents how a ML model is performing. SHAP applies an additive method of feature attribution, i.e., the output model is a linear sum of the input variables, to generate a model that can be interpreted. Considering Eq. 12 that we have a model $x = (x_1, x_2, \ldots, x_P)$ where P is the number of input variables, $g(x')$ is the explanation model with simplified input x' of an original model $f(x)$:

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x'}) = \boldsymbol{\varphi}_0 + \sum_{\boldsymbol{i}=1}^{\boldsymbol{M}} \boldsymbol{\varphi}_{\boldsymbol{i}} \boldsymbol{x_{i'}} \tag{12}$$

Where the M is the number of input features and φ_0 is the constant value representing the model output when no features are present. There is a mapping function between the inputs x and x'. SHAP can therefore provide meaningful insights into local and global model behavior. In order to estimate SHAP values, a number of approaches have been suggested, such as Kernel SHAP, Deep SHAP, and Tree SHAP [85]. In the current study, a model-agnostic Kernel SHAP method was employed, as it is suitable for a wide range of ML architectures, including stacking ensemble and single ML methods [46–48].

3 Results and discussions

3.1 Statistical overview

Statistical analysis, at first glance, is crucial for the current investigation. The dataset includes 84 sets of grinding process data that have one set of numerical which is the removal rate, and four coolant types and three grinding wheel types, both considered qualitative variables associated with the grinding process. At the end of the process, the surface roughness Rz is measured and reported. Table 5 describes an overview of all the included data during the process. However, implementing any statistical and even analytical procedure like ANOVA in qualitative data is impossible, therefore converting qualitative data to quantitative representation is necessary, for current study four different coolant includes (dry, MQL1, MQL2 and fluid) [encoded as integers 1 through 4, respectively, consequently for grinding wheel, three types of different grinding wheel associated include (denoted A, B, C) encoded as integers 1 through 3, respectively. The conversion of qualitative parameters (such as coolant type and grinding wheel type) into quantitative labels is a necessary preprocessing step for ML implementation. Most predictive algorithms

Table 5 Statistical analysis overview of preprocessed dataset

		Min	Max	Average (Mean)	STD
Inputs	Removal Rate (mm)	7.5	97.5	52.5	30.2
	Coolant	1	4	2.5	1.1
	Grinding Wheel	1	3	2	0.8
Outputs	Surface Roughness (μm)	0.4	4.5	1.6	0.8

(e.g., GPR, ANN, and XGBoost) require numerical inputs, and therefore, categorical variables must be encoded to enable effective training. This approach is widely recognized and has been incorporated in several scholarly studies [86–88].

In Table 5, a primary statistical analysis on whole dataset was conducted, it is crucial to note that for any future investigation like ML algorithms the input data represent or known as predictors should be normalized or scaled to a consistent range, while the predicted output accuracy and error measured by mean value of real predictors like (Rz).

3.2 ANOVA

In the ANOVA for many applications, only the sum of squares and p-values are useful; the others are merely computational steps along the way to the p-values that indicate the results of significance testing. Therefore, the p-value is the final result that should be considered. Nonetheless, it is traditional to report all the computed data in a table to substantiate claims of statistical significance in ANOVA. Each hypothesis is tested using an F-test, which compares the mean square for that source of variation (which increases when the source of variation contributes meaningfully to the measured response) to the error mean square. To find out whether or not that source of variation) has a significant effect, simply look at its p-value (p) and decide "significant" if it is small enough, for example, if $p < 0.05$ [89].

Table 6 represents the surface roughness's ANOVA; the results related to different response variables. It is completely clear that all the mentioned predictors (input values) have significant effects on responses because the p-value for all of them lower than 5%, therefore the ANOVA test claims that considering new parameters like types of coolant and grinding wheel will have significant effects on output variables in every mentioned test; However the ANOVA did not tell about how much effect each variables will be have, therefore another approaches like formula implementation and ML implementation is necessary to provided a systematic flow to measure the effects of each parameters on the output value and create a platform to predict new response based on new predictors.

Table 6 ANOVA analysis results for surface roughness (Rz)

Parameter	Sum Sq.	d.f.	Mean Sq.	F	*P*-Value
Removal rate	7.3	6	1.2	6.2	$2.6e^{-5} < 5\%$
Coolant	7.8	3	2.6	13.3	$5.6e^{-7} < 5\%$
Grinding wheel	21.0	2	10.5	53.5	$5.8e^{-15} < 5\%$
Error	14.1	72	0.2		
Total	50.3	83			

Table 7 Analysis of accuracy matrices for formulas evaluated from PSO

Metric	Rz
Mean Accuracy (%)	-2522.76%
RMSE	41.83
MAPE (%)	2622.76%
R^2	-2922

3.3 Particle swarm optimization

PSO is an iterative algorithm that starts with an initial value to reach an optimized solution with respect to minimizing a predefined cost function; for the current implementation, the cost function is mean squared error (MSE), which is defined in Eq. 4. Therefore, the PSO iterates to reach the minimum threshold defined for the cost function value. The current implementation iterates and reaches the threshold value below 0.001 after 20 iterations. The final simplified formula for each response is defined in Eq. 13 for surface roughness. The accuracy metrics are also tabulated in Table 7. From the obtained equation, it is clear that none of the input variables has an insignificant effect on the output response. Their coefficients are large enough that they cannot be neglected, and even the interactions between variables are important and cannot be ignored.

From Table 7, the R^2 for surface roughness is unacceptable, with MAPE exceeding 15%, which shows that working with this equation needs more consideration; however, the equation clearly defines a direct analytical approach for expressing the relation between removal rate, types of coolant, and types of grinding wheel. Additional metric analysis reinforces this concern: the RMSE reaches 41.83, and the MAE is 34.36, both indicating large absolute errors, while the mean accuracy is negative (−2522.76%), revealing systematic bias; the highly negative R^2 further proves that the model performs far worse than simply predicting the mean. Therefore, recalibration or an alternative modelling strategy is strongly recommended. At this stage, further analysis using ML methods is not justified.

$$\begin{aligned} \boldsymbol{R_z} = &-4.27\boldsymbol{x_r} - 5.22\boldsymbol{x_c} - 0.63\boldsymbol{x_w} + \boldsymbol{x_r}^{-4.32} + \\ &\boldsymbol{x_c}^{-8.23} + \boldsymbol{x_w}^{-0.58} - \tfrac{2.85}{\boldsymbol{x_r}} + \tfrac{3.99}{\boldsymbol{x_c}} - \tfrac{8.75}{\boldsymbol{x_w}} + 1.41\boldsymbol{x_r}\boldsymbol{x_c} + \\ &6.55\boldsymbol{x_c}\boldsymbol{x_w} - 0.78\boldsymbol{x_r}\boldsymbol{x_w} + 5.67\tfrac{\boldsymbol{x_r}}{\boldsymbol{x_c}} + 0.84\tfrac{\boldsymbol{x_c}}{\boldsymbol{x_w}} \\ &-1.53\tfrac{\boldsymbol{x_r}}{\boldsymbol{x_w}} - \tfrac{1.27}{\boldsymbol{x_r}\boldsymbol{x_c}} - \tfrac{0.24}{\boldsymbol{x_x}\boldsymbol{x_w}} - \tfrac{0.69}{\boldsymbol{x_r}\boldsymbol{x_w}} \end{aligned} \tag{13}$$

3.4 Machine learning implementation

3.4.1 Single algorithms

Based on model performance, hyperparameters were chosen for each base and meta-model algorithm. Using Bayesian optimization through BayesSearchCV, a technique from the scikit-optimize library, hyperparameter tuning was carried out. By creating a probabilistic model of the objective function and choosing promising hyperparameter combinations, this method effectively explores the hyperparameter space. The size of the training and testing implementation is 68 sets of training data and 16 sets of testing data. Cross-validation was used to evaluate each combination's performance to ensure robust selection of optimal parameters. The optimal hyperparameters for each model are displayed in Tables 8 and 9, and 10. The model with the lowest RMSE, MAPE, and the highest accuracy and R^2 score is the best-performing one.

Figure 6 compares the actual and predicted surface roughness for the three single ML models and shows how well the results fit the regression line or reference line in the scatter plot. For the training dataset, red square markers are used, and for the test data, purple crosses are used. Actual Rz is on the x-axis, and Predicted Rz is on the y-axis.

Each model achieves comparable training accuracy (roughly 97%) and high R2 (0.99), indicating a strong fit. XGBoost has the lowest RMSE (0.03 µm) and MAPE (2.20%) for the training data. Based on the test results, GPR consistently outperforms XGBoost and ANN in surface roughness prediction, as demonstrated by the performance metrics Accuracy, RMSE, MAPE, and R^2 (Table 11). GPR had the lowest RMSE (0.12 µm) and MAPE (7.03%) on the test data. This advantage holds on the test set, where GPR attains 92.96% accuracy and a 0.97 R^2. So, these results mean that its predictions were more stable. XGBoost and ANN result in higher MAPE errors of 10.25% and 9.48%, respectively. These results confirm GPR's strong generalisation capability and its resistance to overfitting. This is an enduring challenge in regression work.

According to the metrics for all data, it is evident that XGBoost and GPR, more than ANN, are the best-suited ML methods in this study. Although ANN is slightly inaccurate, it is still a reliable method. For all data, GPR had an R^2 of 0.99, an accuracy of 97.50%, an RMSE of 0.05 µm, and a MAPE of 2.49%, which means that its predictions were more accurate and stable. Overall, GPR gave reliable results supported by the strong R^2 value.

Therefore, according to the predicted results for the test data, it is clear that the GPR algorithm is better than the other methods. This shows that GPR is strong and can be used with any data split, as detailed in Table 11. The overall outcomes of the XGBoost, GPR, and ANN approaches are shown in Fig. 6a and b, and 6c, respectively.

Table 8 Optimal hyperparameters for XGBoost

Hyperparameter	Optimal value
Number of trees	200
Learning rate/tree	0.1
Maximum depth/tree	5
Subsample/tree	80%
Column sample/tree	80%

Table 9 Optimal hyperparameters for ANN

Hyperparameter	Optimal value
Hidden layers	2 (32, 64)
Node per hidden layer	10
Weight initializer	Random_normal
Activation function	ReLU
Loss function	MSE
Back propagation function	Adam

Table 10 Optimal hyperparameters for GPR

Hyperparameter	Optimal value
Kernel Function	Constant * RBF + WhiteKernel
Constant Kernel Initial Value	1.0
RBF Length Scale	1.0
Normalize Target (normalize_y)	False

3.4.2 Stacking ensemble

In this study, there are three stacking ensemble models: GPR-XGBoost: ANN, GPR-ANN: XGBoost, and ANN-XGBoost: GPR. Table 12 displays the accuracy performance of each of the three stacking ensemble models. The three stacking ensemble models took 70% of the data to train the base model, 15% to train the meta model, and the remaining 15% to test the models. The models were checked by analysing their prediction accuracy metrics.

The training accuracy values were high in all three stacking ensemble models, implying that the models achieved strong fits on the training data. Among them, the ANN-XGBoost: GPR combination performed slightly better than the others, reaching a training accuracy of 97.46%, a MAPE of 2.53% and a low RMSE of 0.06 µm. Although this indicates effective learning on the training data, the gap between training and test performance highlights the need to carefully evaluate generalization. Regarding the meta set, which plays a crucial role in training the meta-learner, the ANN-XGBoost: GPR and GPR-XGBoost: ANN showed the most stable results, with the lowest RMSE and MAPE. On the contrary, GPR-ANN: XGBoost had the poorest result, having an R^2 of 0.81, an RMSE of 0.21 µm, and a MAPE of 16.90%, making it unstable in case of model integration. All three models had good generalization on the test set, which

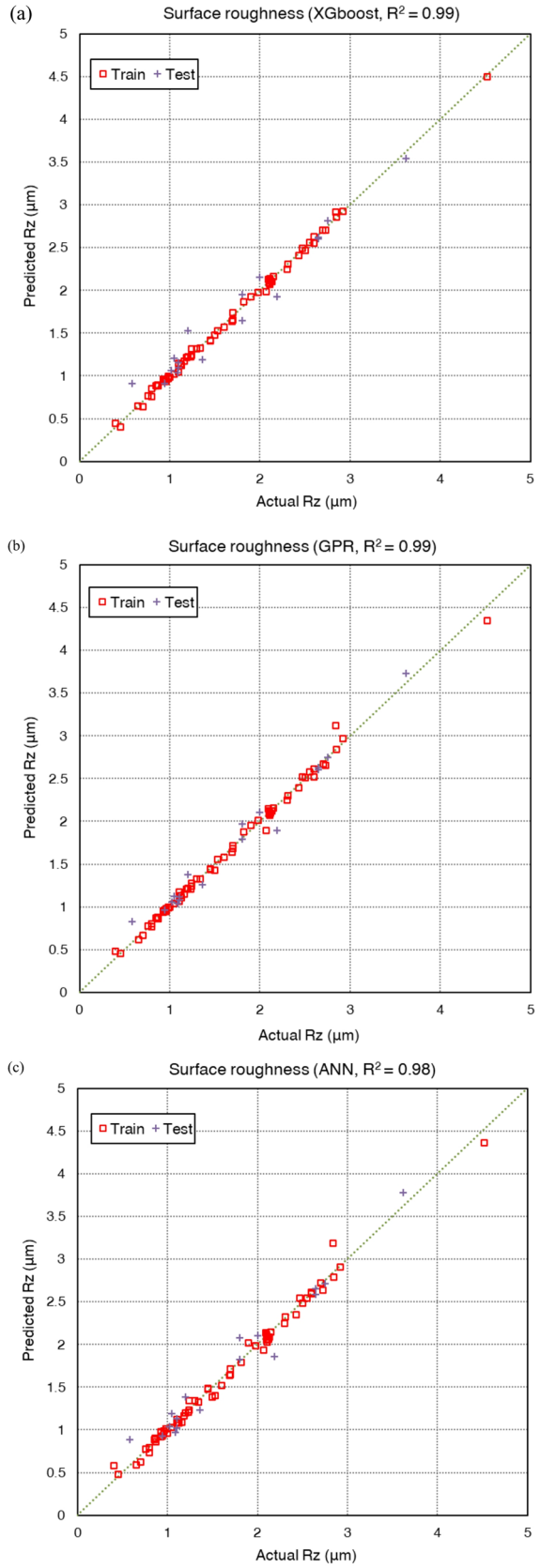

Fig. 6 Correlation between prediction and Actual Rz with ML model: **a** XGBoost; **b** GPR; **c** ANN

Table 11 Accuracy metrics of the single algorithm

Model		Accuracy (%)	RMSE (µm)	MAPE (%)	R^2
XGBoost	Train	97.79	0.03	2.20	0.99
GPR		97.73	0.05	2.26	0.99
ANN		96.25	0.07	3.74	0.99
XGBoost	Test	89.74	0.15	10.25	0.96
GPR		92.96	0.12	7.03	0.97
ANN		90.51	0.15	9.48	0.96
XGBoost	All	97.27	0.04	2.72	0.99
GPR		97.50	0.05	2.49	0.99
ANN		95.56	0.07	4.43	0.98

Table 12 Accuracy metrics of the three stacking ensemble models

Model		Accuracy (%)	RMSE (µm)	MAPE (%)	R^2
GPR-XGBoost: ANN	Train	97.21	0.06	2.78	0.99
GPR-ANN: XGBoost		94.13	0.14	5.87	0.96
ANN-XGBoost: GPR		97.46	0.06	2.53	0.99
GPR-XGBoost: ANN	Meta	85.58	0.17	14.41	0.88
GPR-ANN: XGBoost		83.10	0.21	16.90	0.81
ANN-XGBoost: GPR		85.96	0.15	14.03	0.90
GPR-XGBoost: ANN	Test	89.45	0.16	10.54	0.94
GPR-ANN: XGBoost		89.41	0.16	10.58	0.94
ANN-XGBoost: GPR		90.13	0.15	9.86	0.94
GPR-XGBoost: ANN	All	94.21	0.11	5.78	0.97
GPR-ANN: XGBoost		91.69	0.16	8.30	0.96
ANN-XGBoost: GPR		94.54	0.10	5.45	0.98

measures predictive performance on unseen data. Among them, the ANN-XGBoost: GPR reached a test accuracy of about 90.13% and the lowest MAPE of 9.86%. All test sets had an R^2 of 0.94. These results highlight and emphasize the accurate and strong predictive performance of the stacking ensemble technique when multiple ML methods are combined.

The ANN-XGBoost: GPR model, among the three stacking ensemble models on all the data, showed the best and most balanced measures, due to achieving an R^2 of 0.98, which is an important metric in ML evaluation. The second-best model was GPR-XGBoost: ANN, with an R^2 of 0.9. These two models had high accuracy (94.21% to 94.54%), low RMSE (0.11 and 0.10 µm), and low MAPE (5.78 and 5.45%), meaning perfect generalization ability. GPR-ANN: XGBoost also validated well ($R^2 = 0.96$). However, its overall accuracy was lower (91.69%), and the RMSE (0.16 µm) and MAPE (8.30%) were slightly higher than the other methods. It is concluded that the combination of XGBoost with other single algorithms, using ANN or GPR as the meta base model, was the best at accurately and consistently predicting surface roughness across different datasets. Overall, ANN-XGBoost: GPR is the best among these stacking ensemble methods. In Fig. 7, the predicted and actual surface roughness values are illustrated for each model.

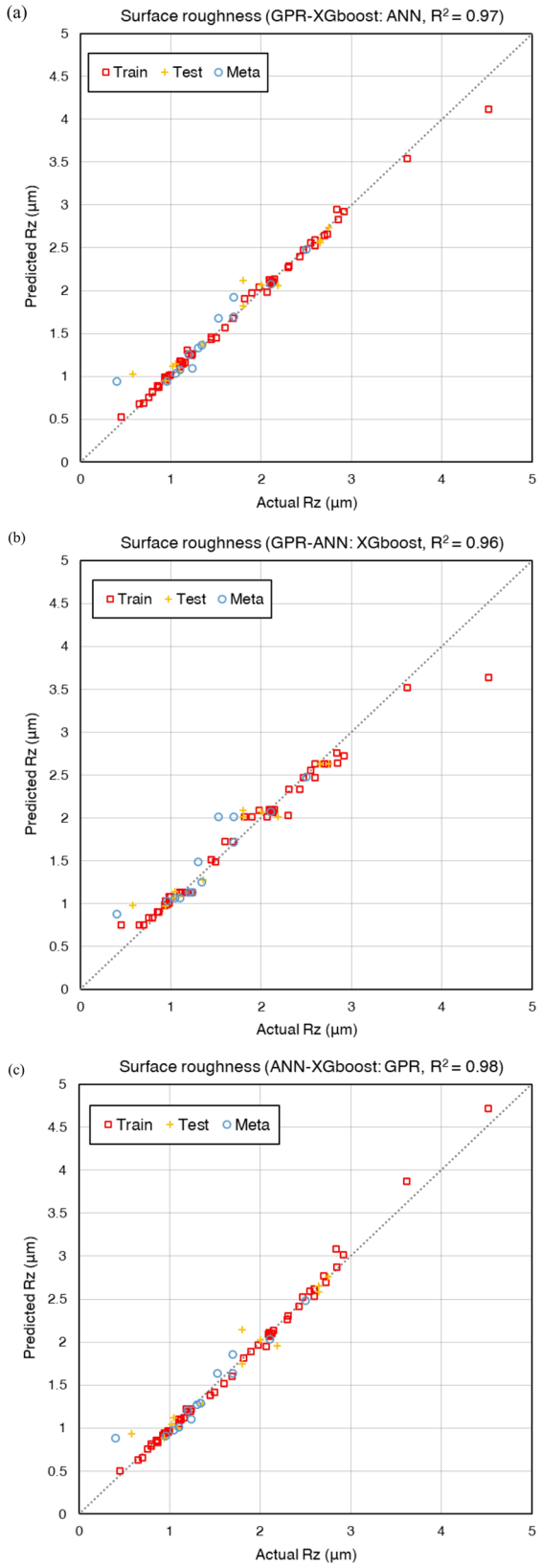

Fig. 7 Correlation between prediction and Actual Rz with Stacking ensemble model: **a** GPR-XGB: ANN; **b** GPR-ANN: XGB; **c** ANN-XGB: GPR

According the accuracy mertics from single models, the GPR method consistently has the best predictive accuracy, with the lowest RMSE and MAPE across all data evaluations. This shows that it is a very stable model that does not overfit. The stacking models, however, perform better at generalizing across various data distributions. This makes them more robust for complex manufacturing datasets. According to prior studies, GPR is most effective in controlled environments with dissimilar datasets [52], but the stacking ensemble method improves model robustness by combining the strengths of different base learners. ANN-XGB: GPR has little higher R^2 and accuracy and slightly lower RMSE and MAPE on the test set than the other stacking models. This suggests that it could be an ideal framework for future predictive modeling. These results show that research should focus on improving stacking architectures by adding advanced meta-learning strategies and domain adaptation techniques. This will make the models more scalable and resilient in environments with vaious variability [90, 91].

The meta learner, which uses the predictions from the base models, makes the accuracy metrics of the stacking models a little lower. For instance, R^2 is reduced by one or three% in the ensemble methods, in comparison with single ML methods. During the testing phase, the smaller amount of data in the stacking models makes it harder to perform a full evaluation, which can lead to a small reduction in prediction performance. The identified disparity can not solely be explained by the limited number of input variables, including the type of coolant, grinding wheel type, and removal rate, which may not fully capture the multifaceted nature of the factors that affect machining performance. The 84 observations used in this study are adequate, as supported by the resulting accuracy metrics. This challenge is common in experimental machining research. Overall, these factors are more likely to make the stacking ensemble slightly less accurate; however, the advantage of this method is that it allows for more general use by reflecting associations with other underlying processes and ML methods [92].

3.5 Sensitivity analysis

Figure 8 presents the sensitivity of surface roughness (Rz) to variations in removal rate and coolant type, evaluated using individual ML models across three grinding wheels denoted A, B, and C, as listed in Table 3. The outcomes demonstrate that there is a steady rise in the surface roughness with an increase in the removal rate, irrespective of the coolant or the grinding wheel.

The fluid type of coolant always has the lowest Rz average values, and the dry condition has the highest roughness. This is because of the fact that, in dry grinding, there is no

lubrication and cooling, leading to more friction and thermal stress. The effect of the grinding wheels is clear and also reflected in Figs. 8 and 9. Type C grinding wheel results in the lowest overall Rz at all removal rates, and especially at low removal rates. B performed better compared to A, which consistently yields the highest values of roughness.

Figure 9 is concentrated on the prediction of the stacking ensemble ML model on the surface roughness prediction with a similar set of input parameters. This ensemble procedure can be used to capture non-linear relationships between the input variables and Rz, and the model is more accurate than single models. Although the removal rate is the most decisive factor in measuring surface roughness, the selection of the grinding wheel is becoming increasingly significant as the removal rate decreases. Once more, the C wheel shows itself to be very high on surface finish performance regardless of coolant type, and this is a further testament to the effectiveness of the Wheel_C in precision grinding applications. The variations between the coolant types, however, diminish at the higher removal rates. Overall, the analysis proves the fact that the removal rate is among the most considerable aspects influencing the surface roughness, whereas the choice of grinding wheel and the type of coolant come next. Fluid coolant, particularly moderate to low removal rates, improves the quality of the surface finish. C is the best of the grinding wheels. These results confirm the capacity of the stacking ensemble model to reflect complicated process interactions and propose that a combination of coolant strategy and grinding wheel selection can considerably decrease the surface roughness in the grinding process.

3.6 Shapley additive explanations

Figure 10 shows SHAP summary plots for the GPR using the test results and the stacking model with all data (84 observations). The y-axis follows input variables, which include two categorical factors (one-hot encoded) and one quantitative factor, the specific removal rate. Each dot represents an individual experiment. The dots on the x-axis show the SHAP value, which is based on how far the prediction shifts from the expected (baseline) surface-roughness value. The colour of the dots represents the corresponding value of that variable (red = low, black = high). A SHAP value demonstrates the distance a single feature moves away from the model prediction towards the baseline (the mean prediction). A SHAP value greater than zero increases the output value, and a SHAP value less than zero decreases it. Using all SHAP values of the features for a given sample explains the difference between that sample's prediction and the model's baseline value. The features are ranked in descending order of their mean absolute SHAP values, so the most or the least influential features in the model can be clearly identified. To interpret the model, SHAP was used on the ANN-XGB: GPR (Fig. 10a) and the single GPR model (Fig. 10b) with the test data. However, the stacking model was used on the entire dataset because of the small size of the test data and potential outliers. This method improves interpretability and demonstrates the efficacy of SHAP on various models and analytical techniques.

According to Fig. 10, the two plots indicate that surface roughness prediction is dominated by grinding wheel type. The analysis shows that the two types of Al_2O_3 wheels have a similar impact on the output. Among the grinding wheel types, 89A180K6V111, A (top row, which gives positive SHAP) is plainly the most influential, Wheel C is the second most influential, and B is the least influential. The black dots of A in the summary plot are far to the right of zero, indicating large positive SHAP values. Therefore, the selection of this wheel always increases surface roughness. Wheel C has the opposite effect to Wheel A. The latter always tends to push the roughness upwards, whereas C tends to pull the roughness down.

The order of influence of coolants is Fluid < MQL2 < Dry < MQL1: Fluid has the most negative SHAP values and consequently reduces the predicted surface roughness the most; it has the highest penetration and wide coverage of the grinding zone, which improves heat removal and lubrication. MQL2 continues to reduce roughness, albeit to a lesser degree; Dry machining swings the SHAP slightly to the positive side, having the tendency to increase roughness; and MQL1 indicates the smallest, almost zero SHAP magnitudes, which makes it the weakest and least consistent coolant in this work. In the Dry coolant, black dots (Dry = 1) are mainly to the right of the zero line, so that the exclusion of coolant moves the predicted surface roughness upwards. Red dots (Dry = 0, any other coolant in use) go to the left, with some minor reduction of roughness. These dots are spread narrowly, and their mean absolute SHAP value is lower than the Fluid or MQL2; thus, the influence of Dry is less and more consistent compared to the other coolant options (Fig. 10a). In the GPR model, the Dry coolant position is substituted by MQL2 in Fig. 10b.

Figure 11 is a SHAP dependence plot of the specific removal rate. The x-axis contains the levels, and the y-axis shows SHAP estimations, and the red-to-black gradient is used in order to visualize the magnitude of the rate itself. This figure examines the only quantitative input, the removal rate, and shows how each discrete level influences the predicted roughness. Because wheel type and coolant are set at the only level that is active in the experiment and are one-hot encoded (one column is 1 and the rest are 0), the individual positive or negative effects of these categorical inputs was presented in Fig. 10. The binary variables

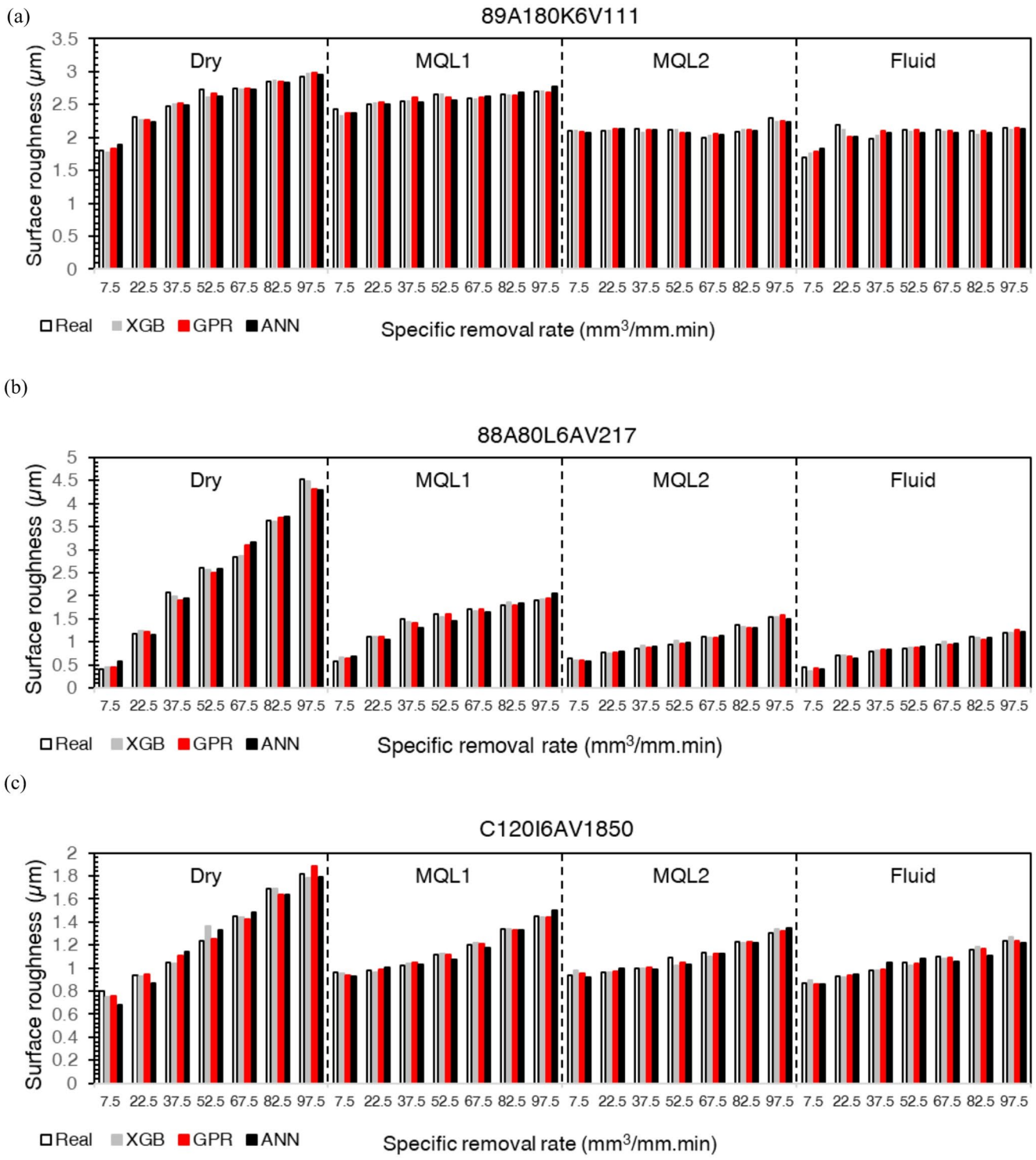

Fig. 8 Analysis of the effect of input parameters on the surface roughness under different removal rates and coolants for the proposed ML models, with different Grinding wheel types: **a** 89A180K6V111; **b** 88A80L6AV217; **c** C120I6AV1850

(one-hot encoded parameters) do not vary within this graph and their impact is not visible here; instead, it is represented in the summary plot (Fig. 10), where each row indicates how the model reacts to the presence (1) or absence (0) of those categorical options.

The most influential continuous factor is the specific removal rate. Figure 11 indicates that SHAP monotonically increases from about −1.0 at 7.5 to +1.2 at 97.5, indicating that higher rates lead to higher roughness. Specific removal rate is the only quantitative input in the study, and it shows

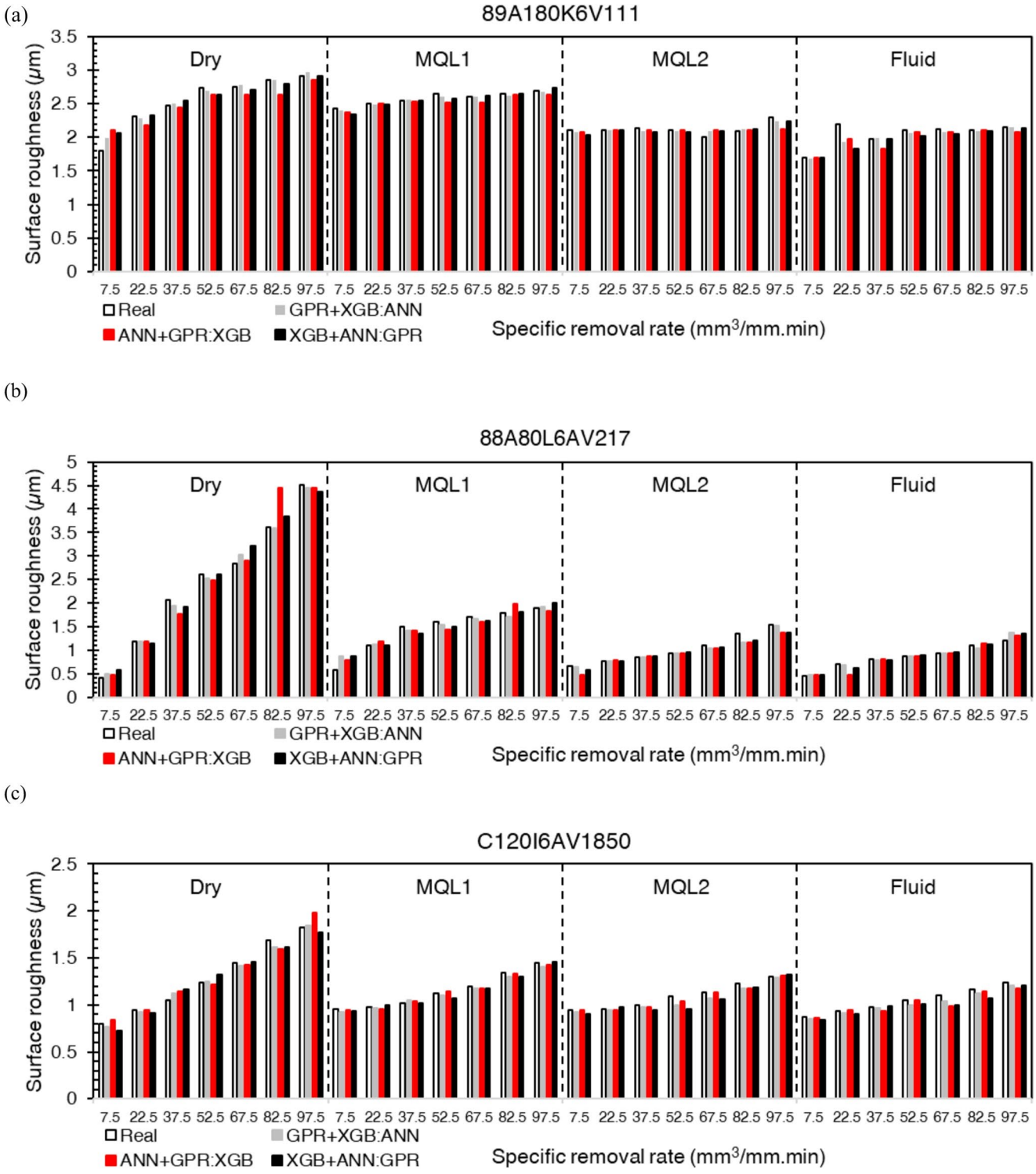

Fig. 9 Analysis of the effect of input parameters on the surface roughness under different removal rates and coolants for the proposed stacking ensemble models, with different Grinding wheel types: **a** 89A180K6V111; **b** 88A80L6AV217; **c** C120I6AV1850

the widest SHAP range and one of the greatest impacts on surface roughness. It has the widest SHAP distribution, with the highest and the lowest individual SHAP values to surface roughness, even though it ranked second in mean absolute importance. Figure 11a is the stacking model that has a roughly linear trend, and the SHAP value reached its maximum at the highest removal rate. Figure 11b, which is related to the GPR model, shows a similar trend, but the

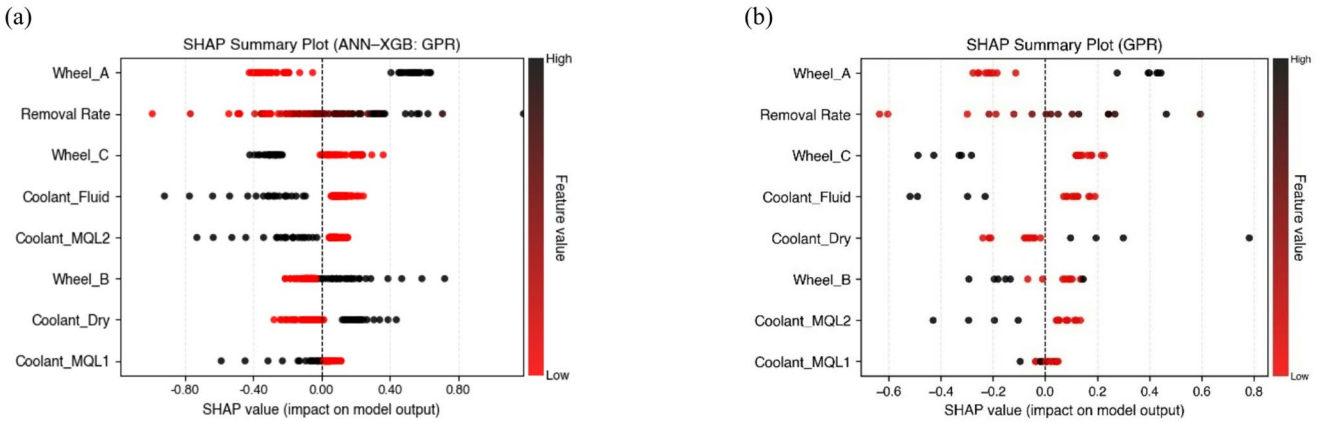

Fig. 10 SHAP summary plot for: **a** ANN-XGB: GPR; **b** GPR models

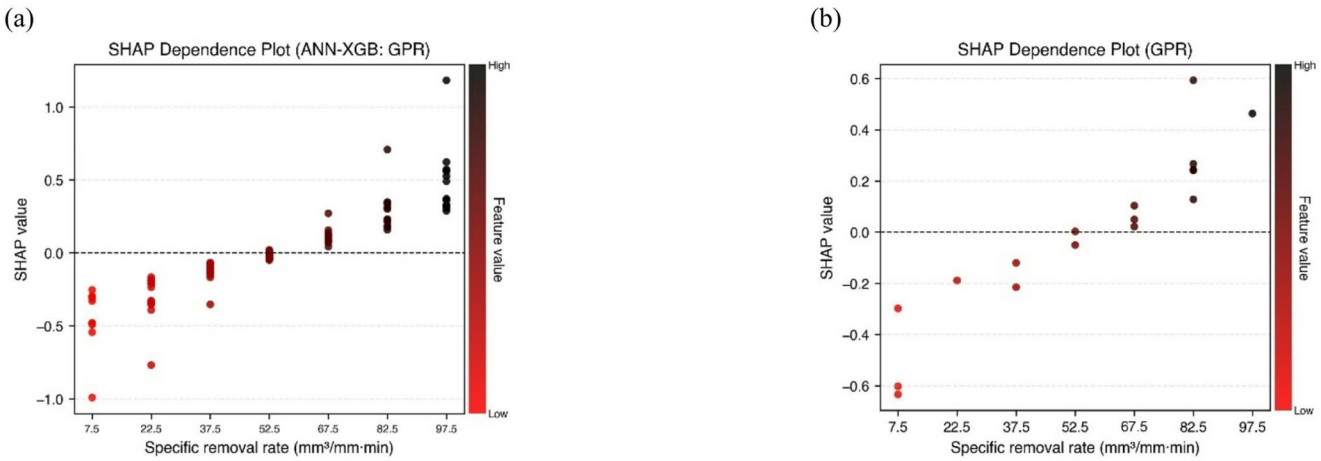

Fig. 11 SHAP dependence plot for: **a** ANN-XGB: GPR; **b** GPR models

maximum SHAP value is at the second-highest removal rate.

The physical and mechanical behaviour of the grinding process was clarified through SHAP analysis with the assistance of sensitivity analysis. It was observed that the grain size of the grinding wheel has a major effect: higher roughness of the surface was related to coarser grain structures. This is the reason why the roughness values were highest with wheel_A and Wheel_B, which are characterised by relatively large grain, whereas Wheel_C, which has the smallest grain size, produced the smoothest surfaces. The specific removal rate turned out to be the most significant factor in general; an increase in this rate was consistently followed by a rise in surface roughness, as it is in line with the trends of Figs. 8 and 9. Conversely, the coolant type had the lowest mean SHAP contribution. This result was expected, since coolants do not directly increase the surface roughness, but are used to ensure the surface integrity by limiting thermal damage, burning of the surface, and the formation of residual stresses during and after grinding.

The results of the sensitivity analysis showed that the fluid coolant had the most favourable effect, and the dry condition had the weakest performance. The SHAP results also showed that dry grinding had more severe effects on the surface than the MQL conditions (Fig. 10a). In the MQL group, MQL1 was predicted to have the least influence. This can be attributed to its relatively high viscosity, which limits its ability to spread over the workpiece, and thus reduces its ability to cool and lubrication performance compared with vegetable oil and water-miscible alternatives. The stacking model reproduced these physical relationships in a more understandable way by properly identifying that dry grinding is the condition most detrimental to surface quality. Overall, SHAP with sensitivity analysis proves that the stacking ensemble model is effective in identifying and modeling these physical mechanisms.

4 Conclusion

This work is an explainable prediction framework for surface roughness in grinding. This study compared a collection of data-based methods with an 84-run dataset from grinding experiments on AA6061 aluminum alloy using an M7135A NANTONG SHUANGZANG-NC surface grinding machine in down-grinding mode. In this study, single ML models, stacking ensemble algorithms, sensitivity analysis, and SHAP interpretation were integrated to explain how the specific removal rate, coolant type, and grinding-wheel type dominate surface roughness.

Conventional equation-based fitting was not reliable. This motivates the use of machine-learning methods. GPR, ANN, and XGBoost were the ML algorithms that were applied to determine the relationship between input variables and the output variable. The GPR model yielded the best single performance with an accuracy of 97.50%, an RMSE of 0.05 µm, an MAPE of 2.49%, and R^2 of 0.99. A stacking ensemble model that had two base-learner ML models and one meta-learner ML model was implemented, combining the single ML models. The ensemble with ANN and XGB as base learners and GPR as the meta-learner gave the best trade-off between bias and variance and resulted in an overall accuracy of 94.54%, a RMSE of 0.10 µm, a MAPE of 5.45%, and an R^2 of 0.98. This makes ensemble models slightly less precise but more generalizable, as they reflect associations with underlying process dynamics.

Sensitivity analysis indicate that the Surface roughness increases steadily with the removal rate in any coolant strategy and in any grinding wheel condition. Fluid coolant provides the best Rz, and dry grinding causes the roughest surface due to additional friction and thermal damage. The choice of wheels remains a critical factor. C120I6AV1850 produces the smoothest finish, 88A80L6AV217 yields a moderately rough surface, and 89A180K6V111 results in the surface roughness, especially at higher removal rates. The stacking ensemble model is the one that most accurately captures these nonlinear interactions compared to single models. Thus, the most effective path to reducing surface roughness is combining fluid coolant with the C120I6AV1850 wheel at low specific removal rates.

In the last section, Kernel SHAP, applicable to any ML architecture, was used. The results of SHAP analysis made clear that grinding wheel type 89A180K6V111 was the most significant input, which increased the surface roughness. MQL1 had the least impact, close to zero. Thus, it was the weakest and lowest effect in this work. The specific removal rate was the most significant quantitative input factor. In the case of the removal rate, the SHAP range was the widest, with approximately minus 1 at 7.5 (mm^3/mm.min) and 1.2 at 97.5 (mm3/mm.min), and its analysis demonstrated that increasing the removal rate leads to increased predicted roughness. Overall, the SHAP value increases with a higher removal rate.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Ruzova TA, Haddadi B (2025) Surface roughness and its measurement methods - Analytical review. Results Surf Interfaces 19:100441. https://doi.org/10.1016/j.rsurfi.2025.100441
2. Sushil K, Ramkumar J, Chandraprakash C (2025) Surface roughness analysis: A comprehensive review of measurement techniques, methodologies, and modeling. J Micromanufacturing 25165984241305225. https://doi.org/10.1177/25165984241305225
3. He T, Chen W, Liu Z et al (2025) The impact of surface roughness on the friction and wear performance of GCr15 bearing steel. Lubricants 13. https://doi.org/10.3390/lubricants13040187
4. Chen C, Bai Q, Zhao C et al (2024) Effect of ultrasonic high-frequency micro-forging on the wear resistance of a Fe-base alloy coating deposited by high-speed laser cladding process. Vacuum 221:112934. https://doi.org/10.1016/j.vacuum.2023.112934
5. Jiang A, Zhao J, Cui P et al (2024) Effects of TiAlN Coating thickness on machined surface roughness, surface residual stresses, and fatigue life in turning inconel 718. https://doi.org/10.3390/met14080940. Metals 14:
6. Prasad J, Sonwani RK (2025) Optimize chemical milling of aluminium alloys to achieve minimum surface roughness in Aerospace and Defense Industry. J Indian Chem Soc 102:101537. https://doi.org/10.1016/j.jics.2024.101537
7. Gruber K, Smolina I, Dziedzic R et al (2024) Tailoring heat treatment for AlSi7Mg0.6 parts with as-built surface generated by laser powder bed fusion to reduce surface roughness sensitivity. J Alloys Compd 984:173903. https://doi.org/10.1016/j.jallcom.2024.173903

8. Abellán-Nebot JV, Vila Pastor C, Siller HR (2024) A review of the factors influencing surface roughness in machining and their impact on sustainability. Sustainability 16. https://doi.org/10.3390/su16051917
9. Chen J, Fang Q, Li P (2015) Effect of grinding wheel spindle vibration on surface roughness and subsurface damage in brittle material grinding. Int J Mach Tools Manuf 91:12–23. https://doi.org/10.1016/j.ijmachtools.2015.01.003
10. Yashwant Bhise V, Jogi BF (2022) Effect of cutting speed and feed on surface roughness in dry turning of Inconel X-750. Materials Today: Proceedings 61:587–592. https://doi.org/10.1016/j.matpr.2022.04.098
11. McDonald A, Mohamed A-MO, Warkentin A, Bauer RJ (2017) Kinematic simulation of the uncut chip thickness and surface finish using a reduced set of 3D grinding wheel measurements. Precis Eng 49:169–178. https://doi.org/10.1016/j.precisioneng.2017.02.005
12. Lopes WN, Aguiar PR, Dotto FRL et al (2021) Method for fault detection of aluminum oxide grinding wheel cutting surfaces using a piezoelectric diaphragm and digital signal processing techniques. Measurement 180:109503. https://doi.org/10.1016/j.measurement.2021.109503
13. Barmouz M, Azarhoushang B (2025) Grinding Performance Evaluation of Additively Manufactured Vitrified Bond Grinding Wheel: Tool Wear, Grinding Force, Surface Roughness, and Surface Topography Analysis. Int J Precision Eng Manufacturing-Green Technol. https://doi.org/10.1007/s40684-024-00684-y
14. Novoselov Yu, Bratan S, Bogutsky V (2016) Analysis of Relation between Grinding Wheel Wear and Abrasive Grains Wear. Procedia Eng 150:809–814. https://doi.org/10.1016/j.proeng.2016.07.116
15. Umucu Y, Deniz V, Bozkurt V, Fatih Çağlar M (2016) The evaluation of grinding process using artificial neural network. Int J Miner Process 146:46–53. https://doi.org/10.1016/j.minpro.2015.11.013
16. Qazani MRC, Amini S, Pedrammehr S et al (2023) A machine learning method for cutting parameter selection in rotary ultrasonic-assisted end grinding. Int J Adv Manuf Technol 126:1577–1591. https://doi.org/10.1007/s00170-023-11196-5
17. Pandiyan V, Caesarendra W, Tjahjowidodo T, Tan HH (2018) In-process tool condition monitoring in compliant abrasive belt grinding process using support vector machine and genetic algorithm. J Manuf Process 31:199–213. https://doi.org/10.1016/j.jmapro.2017.11.014
18. Modanloo V, Elyasi M, Lee T, Quagliato L (2025) Modeling of tensile strength and wear resistance in friction stir processed MMCs by metaheuristic optimization and supervised learning. Int J Adv Manuf Technol 139:3095–3118. https://doi.org/10.1007/s00170-025-16070-0
19. Arriandiaga A, Portillo E, Sánchez JA et al (2017) Recurrent ANN-based modelling of the dynamic evolution of the surface roughness in grinding. Neural Comput Appl 28:1293–1307. https://doi.org/10.1007/s00521-016-2568-1
20. Pontes FJ, Ferreira JR, Silva MB et al (2010) Artificial neural networks for machining processes surface roughness modeling. Int J Adv Manuf Technol 49:879–902. https://doi.org/10.1007/s00170-009-2456-2
21. Balonji S, Tartibu LK, Okokpujie IP (2023) Prediction analysis of surface roughness of aluminum Al6061 in end milling CNC machine using soft computing Techniques. Appl Sci 13. https://doi.org/10.3390/app13074147
22. Shahani NM, Zheng X, Liu C et al (2021) Developing an XGBoost Regression Model for Predicting Young's Modulus of Intact Sedimentary Rocks for the Stability of Surface and Subsurface Structures. Front Earth Sci Volume. https://doi.org/10.3389/feart.2021.761990. 9-2021
23. Mustapha IB, Abdulkareem Z, Abdulkareem M, Ganiyu A (2024) Predictive modeling of physical and mechanical properties of pervious concrete using XGBoost. Neural Comput Appl 36:9245–9261. https://doi.org/10.1007/s00521-024-09553-w
24. Ou T, Liu J, Liu F et al (2023) Coupling of XGBoost ensemble methods and discrete element modelling in predicting autogenous grinding mill throughput. Powder Technol 422:118480. https://doi.org/10.1016/j.powtec.2023.118480
25. Gao K, Chen H, Zhang X et al (2019) A novel material removal prediction method based on acoustic sensing and ensemble XGBoost learning algorithm for robotic belt grinding of Inconel 718. Int J Adv Manuf Technol 105:217–232. https://doi.org/10.1007/s00170-019-04170-7
26. Samavatian M, Samavatian V, Kanabar B et al (2025) Analytically guided machine learning for surface roughness determination in steel grinding processes. Ironmak Steelmaking 03019233251357108. https://doi.org/10.1177/03019233251357108
27. Manjunath K, Tewary S, Khatri N, Cheng K (2021) Monitoring and Predicting the Surface Generation and Surface Roughness in Ultraprecision Machining: A Critical Review. Machines 9:369. https://doi.org/10.3390/machines9120369
28. Shang S, Wang C, Liang X et al (2023) Surface roughness prediction in ultra-precision milling: An extreme learning machine method with data fusion. Micromachines 14. https://doi.org/10.3390/mi14112016
29. Bai L, Yang Q, Cheng X et al (2023) A hybrid physics-data-driven surface roughness prediction model for ultra-precision machining. Sci China Technological Sci 66:1289–1303. https://doi.org/10.1007/s11431-022-2358-4
30. Ruan P, Saxena D, Cao J et al (2025) NASPrecision: Neural Architecture Search-Driven Multi-Stage Learning for surface roughness prediction in ultra-precision machining. Expert Syst Appl 262:125540. https://doi.org/10.1016/j.eswa.2024.125540
31. Chen B, Zha J, Cai Z, Wu M (2025) Predictive modelling of surface roughness in precision grinding based on hybrid algorithm. CIRP J Manufact Sci Technol 59:1–17. https://doi.org/10.1016/j.cirpj.2025.02.004
32. Lee S, Chen Z, Luo Y et al (2024) Enhanced prediction accuracy in high-speed grinding of brittle materials using advanced machine learning techniques. J Intell Manuf. https://doi.org/10.1007/s10845-024-02532-x
33. Wang D, Han C, Wang L et al (2023) Surface roughness prediction of large shaft grinding via attentional CNN-LSTM fusing multiple process signals. Int J Adv Manuf Technol 126:4925–4936. https://doi.org/10.1007/s00170-023-11454-6
34. Guo W, Wu C, Ding Z, Zhou Q (2021) Prediction of surface roughness based on a hybrid feature selection method and long short-term memory network in grinding. Int J Adv Manuf Technol 112:2853–2871. https://doi.org/10.1007/s00170-020-06523-z
35. Shehzad A, Rui X, Ding Y et al (2024) Deep-learning-assisted online surface roughness monitoring in ultraprecision fly cutting. Sci China Technological Sci 67:1482–1497. https://doi.org/10.1007/s11431-023-2615-4
36. Sizemore NE, Nogueira ML, Greis NP, Davies MA (2022) Application of machine learning for improved surface quality classification in ultra-precision machining of germanium. J Manuf Syst 65:296–316. https://doi.org/10.1016/j.jmsy.2022.09.001
37. Balonji S, Tartibu LK, Okokpujie IP (2023) Prediction Analysis of Surface Roughness of Aluminum Al6061 in End Milling CNC Machine Using Soft Computing Techniques. Appl Sci 13:4147. https://doi.org/10.3390/app13074147
38. Abbas AT, Sharma N, Al-Bahkali EA et al (2023) A Machine Learning Perspective to the Investigation of Surface Integrity of Al/SiC/Gr Composite on EDM. J Manuf Mater Process 7:163. https://doi.org/10.3390/jmmp7050163

39. Gurgenc T, Altay O (2022) Surface roughness prediction of wire electric discharge machining (WEDM)-machined AZ91D magnesium alloy using multilayer perceptron, ensemble neural network, and evolving product-unit neural network. 64:350–362. https://doi.org/10.1515/mt-2021-2034
40. Baraheni M, Soudmand BH, Amini S, Fotouhi M (2024) Stacked generalization ensemble learning strategy for multivariate prediction of delamination and maximum thrust force in composite drilling. J Compos Mater 58:3113–3138. https://doi.org/10.1177/00219983241289494
41. Natarajan E, Ramasamy M, Elango S et al (2025) Ensemble learning-based metamodel for enhanced surface roughness prediction in polymeric machining. Machines 13. https://doi.org/10.3390/machines13070570
42. Jones T, Cao Y (2025) Tool wear prediction based on multisensor data fusion and machine learning. Int J Adv Manuf Technol 137:5213–5225. https://doi.org/10.1007/s00170-025-15472-4
43. Huang W, Jin Q, Guo X, Na B (2025) Interpretative analyses for milling surface roughness prediction in thermally modified timber: Shapley value (SHAP) and local interpretable model-agnostic explanations (LIME). Wood Mater Sci Eng 20:744–752. https://doi.org/10.1080/17480272.2025.2466218
44. Sharma Timilsina M, Sen S, Uprety B et al (2024) Prediction of HHV of fuel by Machine learning Algorithm: Interpretability analysis using Shapley Additive Explanations (SHAP). Fuel 357:129573. https://doi.org/10.1016/j.fuel.2023.129573
45. Pasic M, Marinkovic D, Lukic D et al (2024) Prediction and optimization of surface roughness and cutting forces in turning process using ANN, SHAP analysis, and hybrid MCDM method. Appl Sci 14. https://doi.org/10.3390/app142311386
46. Feng J, Liang J, Qiang Z et al (2023) A hybrid stacked ensemble and Kernel SHAP-based model for intelligent cardiotocography classification and interpretability. BMC Med Inf Decis Mak 23:273. https://doi.org/10.1186/s12911-023-02378-y
47. Pelegrina GD, Duarte LT, Grabisch M (2023) A k-additive Choquet integral-based approach to approximate the SHAP values for local interpretability in machine learning. Artif Intell 325:104014. https://doi.org/10.1016/j.artint.2023.104014
48. Zhang F, Ono N, Kanaya S (2025) Interpret Gaussian Process Models by Using Integrated Gradients. Mol Inf 44:e202400051. https://doi.org/10.1002/minf.202400051
49. Gholami H, Darvishi E, Moradi N et al (2024) An interpretable (explainable) model based on machine learning and SHAP interpretation technique for mapping wind erosion hazard. Environ Sci Pollut Res 31:64628–64643. https://doi.org/10.1007/s11356-024-35521-x
50. Dehghanpour Abyaneh M, Narimani P, Hadad M, Attarsharghi S (2023) Using machine learning and optimization for controlling surface roughness in grinding of St37. Energy Equip Syst 11:321–337
51. Hadad M, Attarsharghi S, Dehghanpour Abyaneh M et al (2024) Exploring New Parameters to Advance Surface Roughness Prediction in Grinding Processes for the Enhancement of Automated Machining. J Manuf Mater Process 8:41
52. Dehghanpour Abyaneh M, Narimani P, Javadi MS et al (2024) Predicting Surface Roughness and Grinding Forces in UNS S34700 Steel Grinding: A Machine Learning and Genetic Algorithm Approach to Coolant Effects. Physchem 4:495–523. https://doi.org/10.3390/physchem4040035
53. Hadad MJ, Tawakoli T, Sadeghi MH, Sadeghi B (2012) Temperature and energy partition in minimum quantity lubrication-MQL grinding process. Int J Mach Tools Manuf 54–55:10–17. https://doi.org/10.1016/j.ijmachtools.2011.11.010
54. Hadad M, Hadi M (2013) An investigation on surface grinding of hardened stainless steel S34700 and aluminum alloy AA6061 using minimum quantity of lubrication (MQL) technique. Int J Adv Manuf Technol 68:2145–2158. https://doi.org/10.1007/s00170-013-4830-3
55. Rouder JN, Engelhardt CR, McCabe S, Morey RD (2016) Model comparison in ANOVA. Psychon Bull Rev 23:1779–1786. https://doi.org/10.3758/s13423-016-1026-5
56. Strobl P, Schermer E, Groetsch D et al (2022) Identification and Validation of Linear Friction Models Using ANOVA and Stepwise Regression. Lubricants 10:286. https://doi.org/10.3390/lubricants10110286
57. Cardinal RN, Aitken MRF (2013) ANOVA for the Behavioral Sciences Researcher, 0 edn. Psychology
58. Algina J, Olejnik S (2003) Conducting Power Analyses for Anova and Ancova in between-Subjects Designs. Eval Health Prof 26:288–314. https://doi.org/10.1177/0163278703255248
59. ANOVA (Analysis of Variance) In: Statistics Solutions. https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/anova/. Accessed 2 Jul 2024
60. Wang D, Tan D, Liu L (2018) Particle swarm optimization algorithm: an overview. Soft Comput 22:387–408. https://doi.org/10.1007/s00500-016-2474-6
61. Asokan P, Baskar N, Babu K et al (2005) Optimization of Surface Grinding Operations Using Particle Swarm Optimization Technique. J Manuf Sci Eng 127:885–892. https://doi.org/10.1115/1.2037085
62. Lee TS, Ting TO, Lin YJ, Htay T (2007) A particle swarm approach for grinding process optimization analysis. Int J Adv Manuf Technol 33:1128–1135. https://doi.org/10.1007/s00170-006-0538-y
63. Schulz E, Speekenbrink M, Krause A (2018) A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. J Math Psychol 85:1–16. https://doi.org/10.1016/j.jmp.2018.03.001
64. Uçar F, Katı N, MACHINE LEARNING BASED PREDICTIVE MODEL FOR SURFACE ROUGHNESS IN CYLINDRICAL GRINDING OF AL BASED METAL MATRIX COMPOSITE (2020) EJT 10:415–430. https://doi.org/10.36222/ejt.773093
65. Alajmi MS, Almeshal AM (2021) Modeling of Cutting Force in the Turning of AISI 4340 Using Gaussian Process Regression Algorithm. Appl Sci 11:4055. https://doi.org/10.3390/app11094055
66. Pan Y, Zeng X, Xu H et al (2021) Evaluation of Gaussian process regression kernel functions for improving groundwater prediction. J Hydrol 603:126960. https://doi.org/10.1016/j.jhydrol.2021.126960
67. Shyamala Devi M, Abhishek Rao N, Kushal Kumar SG et al (2022) Kernel Feature Variant-Based Gaussian Process Regression for Prediction of Snail Rings. In: Pandian AP, Fernando X, Haoxiang W (eds) Computer Networks, Big Data and IoT. Springer Nature Singapore, Singapore, pp 493–504
68. Setiono R, Wee Kheng Leow, Zurada JM (2002) Extraction of rules from artificial neural networks for nonlinear regression. IEEE Trans Neural Netw 13:564–577. https://doi.org/10.1109/TNN.2002.1000125
69. Landi A, Piaggi P, Laurino M, Menicucci D (2010) Artificial neural networks for nonlinear regression and classification. In: 2010 10th International Conference on Intelligent Systems Design and Applications. pp 115–120
70. Bircanoğlu C, Arıca N (2018) A comparison of activation functions in artificial neural networks. In: 2018 26th Signal Processing and Communications Applications Conference (SIU). pp 1–4
71. Arriandiaga A, Portillo E, Sánchez JA et al (2016) A new approach for dynamic modelling of energy consumption in the grinding process using recurrent neural networks. Neural Comput Applic 27:1577–1592. https://doi.org/10.1007/s00521-015-1957-1
72. Wu J, Li Y, Qiao H et al (2023) Prediction of mechanical properties and surface roughness of FGH4095 superalloy treated by

laser shock peening based on XGBoost. J Alloys Metall Syst 1:100001. https://doi.org/10.1016/j.jalmes.2023.100001
73. Wang ZH, Liu YF, Wang T et al (2024) Intelligent prediction model of mechanical properties of ultrathin niobium strips based on XGBoost ensemble learning algorithm. Comput Mater Sci 231:112579. https://doi.org/10.1016/j.commatsci.2023.112579
74. Hajihosseinlou M, Maghsoudi A, Ghezelbash R (2024) Stacking: A novel data-driven ensemble machine learning strategy for prediction and mapping of Pb-Zn prospectivity in Varcheh district, west Iran. Expert Syst Appl 237:121668. https://doi.org/10.1016/j.eswa.2023.121668
75. Breiman L (1996) Stacked regressions. Mach Learn 24:49–64. https://doi.org/10.1007/BF00117832
76. Breiman L (1996) Bagging predictors. Mach Learn 24:123–140. https://doi.org/10.1007/BF00058655
77. Naser M, Alavi A (2020) Insights into performance fitness and error metrics for machine learning. arXiv preprint arXiv 200600887. https://doi.org/10.48550/arXiv.2006.00887
78. Buyrukoğlu S, Savaş S (2023) Stacked-Based Ensemble Machine Learning Model for Positioning Footballer. Arab J Sci Eng 48:1371–1383. https://doi.org/10.1007/s13369-022-06857-8
79. Olasehinde OO, Johnson OV, Olayemi OC (2020) Evaluation Of Selected Meta Learning Algorithms For The Prediction Improvement Of Network Intrusion Detection System. In: 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS). pp 1–7
80. Naser MZ, Alavi AH (2023) Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences. Archit Struct Constr 3:499–517. https://doi.org/10.1007/s44150-021-00015-8
81. Hodson TO (2022) Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geosci Model Dev 15:5481–5487. https://doi.org/10.5194/gmd-15-5481-2022
82. Naser MZ, Alavi A (2023) Insights into Performance Fitness and Error Metrics for Machine Learning. Archit Struct Constr 3:499–517. https://doi.org/10.1007/s44150-021-00015-8
83. Kvalseth TO (1985) Cautionary Note about R 2. Am Stat 39:279. https://doi.org/10.2307/2683704
84. Plevris V, Solorzano G, Bakas NP, Ben Seghier MEA (2022) Investigation of performance metrics in regression analysis and machine learning-based prediction models. https://doi.org/10.23967/eccomas.2022.155
85. Mangalathu S, Hwang S-H, Jeon J-S (2020) Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Eng Struct 219:110927. https://doi.org/10.1016/j.engstruct.2020.110927
86. Prashanth GS, Sekar P, Bontha S, Balan ASS (2023) Grinding parameters prediction under different cooling environments using machine learning techniques. Mater Manuf Processes 38:235–244. https://doi.org/10.1080/10426914.2022.2116043
87. Hu Z, Chen H, Zhang K, Wei H (2025) Model-Based Prediction of Grinding Surface Roughness With Error Correction via a Knowledge-Based Fuzzy Broad Learning System. IEEE Trans Instrum Meas 74:1–14. https://doi.org/10.1109/TIM.2025.3574910
88. Hu Z, Wang T, Chen H et al (2025) IEEE/ASME Trans Mechatron 1–11. Physics-Guided Meta-Learning for Surface Roughness Prediction Under Various Working Conditions With Limited Data. https://doi.org/10.1109/TMECH.2025.3598351
89. Javadi MS, Ehteshamfar MV, Adibi H (2023) A comprehensive analysis and prediction of the effect of groove shape and volume fraction of multi-walled carbon nanotubes on the polymer 3D-printed parts in the friction stir welding process. Polym Test 117:107844. https://doi.org/10.1016/j.polymertesting.2022.107844
90. Wang Q, Lu H (2024) A novel stacking ensemble learner for predicting residual strength of corroded pipelines. npj Mater Degrad 8:87. https://doi.org/10.1038/s41529-024-00508-z
91. Lu M, Hou Q, Qin S et al (2023) A stacking ensemble model of various machine learning models for daily runoff forecasting. Water 15. https://doi.org/10.3390/w15071265
92. Faska Z, Khrissi L, Haddouch K, El Akkad N (2023) A robust and consistent stack generalized ensemble-learning framework for image segmentation. J Eng Appl Sci 70:74. https://doi.org/10.1186/s44147-023-00226-4