

Machine Learning-Based Surface Roughness Prediction In Turning Of Hardened AISI 4340 Steels: Incorporating Tool Wear Via Cutting Length

A. Ginting^{1*}, M.S. Kasim², and B.T. Hang Tuah Baharudin³

¹Laboratory of Machining Processes, Department of Mechanical Engineering, Faculty of Engineering, Universitas Sumatera Utara, Jalan Almamater, J17.01.01, Kampus USU, 20155, Medan, Indonesia

²CADCAM Research Group, Faculty of Innovative Design and Technology, Universiti Sultan Zainal Abidin, Kampus Gong Badak, 23100, Kuala Nerus, Terengganu, Malaysia

³Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400, UPM Selangor, Selangor Darul Ehsan, Malaysia

*Corresponding author. E-mail: armansyah.ginting@usu.ac.id

Received: Jul. 30, 2024; Accepted: Mar. 03, 2025

This study explores the predictive modeling of surface roughness in the hard turning of AISI 4340 steel using machine learning techniques, specifically Random Forest (RF) and Gaussian Process Regression (GPR). The uncoated carbide was used and the experimental design considered varying cutting speeds (70, 90, 110 m/min), feeds (0.14, 0.16, 0.22 mm/rev), depths of cut (0.25, 0.5 mm), and cutting length (80, 160, 240 mm) to account for tool wear as an uncontrollable factor. The RF model achieved an RMSE of $0.1627\mu\text{m}$ and an R^2 value of 0.9718, while the GPR model had an RMSE of $0.1676\mu\text{m}$ and an R^2 value of 0.8333. The novelty of this study lies in considering the influence of tool wear via cutting length, significantly impacting the RMSE of the GPR model. Using K-fold cross-validation (K=7) on a 50% training dataset resulted in the lowest RMSE values for both models. Despite the GPR model's slightly lower performance, it demonstrated robustness and consistency across different cross-validation splits and random states, making it a reliable option for predicting surface roughness. This research provides insights into the application of machine learning for process optimization in hard turning operations, highlighting the importance of tool wear and training dataset size. Future work could extend these findings to other machining processes and material types to validate the models' robustness and generalizability.

Keywords: Cutting length; Gaussian Process Regression; Hard turning; Model performance; Random Forest.

© The Author(s). This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

[http://dx.doi.org/10.6180/jase.202601_29\(1\).0003](http://dx.doi.org/10.6180/jase.202601_29(1).0003)

1. Introduction

Hard machining has been coined and implemented in the metal cutting industry since the 1970s. This technology aims to raise productivity by processing metal directly in its hardened condition. Implementing hard machining requires a rigid machine tool, suitable cutting tool material, and an appropriate cutting condition [1].

Studies related to hard machining are mostly about selecting cutting tool materials and geometry, as well as cut-

ting conditions. From the 1970s to the 1980s, it was reported that carbide, ceramic, and cubic boron nitride (CBN) were the suitable cutting tool materials used for hard machining various hardened steel from 50 HRC to fully hardened [1]. The study within the boundary conditions for hard machining implementation continued, and in [2], the mechanisms of chip removal and the thermo-mechanical influence of the product were reported. For turning hardened steels, in particular, several models of chip removal were introduced. The metallurgical fundamentals, the stress, and the

temperature distributions in hard turning were discussed and summarized. The study also focused on determining hard machining processes' achievable workpiece quality and economic efficiency.

The workpiece quality of machined components used in various structures like transportation (airplanes, ships, trains, cars) is subjected to cyclic and randomly variable loads, which may lead to failure due to fatigue. The fatigue behavior of materials is empirically influenced by surface roughness and residual stress [3]. The study of [4, 5] reported that surface roughness greatly influenced the fatigue limit, where increasing Ra reduced fatigue resistance.

Hard turning is normally carried out under dry conditions [1, 6, 7]. Dry conditions promote the thermal softening of the workpiece material, and when hard turning is used with ceramic tool material, thermal shock can be avoided. However, the absence of cutting fluid when hard turning under dry conditions promotes higher surface roughness. In this case, and considering the issue of surface roughness influencing the fatigue limits, the effort to study surface roughness in hard turning is supported by a minimal quantity of lubricant (MQL) technique. The success of MQL in reducing surface roughness (about 7% to 13%) when turning the hardened stainless steels was reported in [8]. The similar results were also reported in [9, 10]. More comprehensively, the study related to machined surface quality was reported in [6, 11].

With the introduction of machine learning in manufacturing, surface roughness has been studied in the intelligent manufacturing field. Recent advances in machine learning applications in machining processes were comprehensively reviewed and presented in [12]. Machine learning applications were reportedly utilized in several aspects of machining processes, including surface roughness quality prediction and tool condition monitoring. Segreto and Teti [13] used a machine learning approach to determine the state of the polishing process in terms of target surface roughness achievement. The other comprehensive reviews were also reported in [14, 15]. In the papers, the authors underlined six strategic topics to be studied under data-driven smart manufacturing that is mainly applied for machining monitoring systems, and surface roughness was one of the parameters that were urgently monitored.

As discussed earlier, the experimental-based study of surface roughness is potentially enriched with machine learning techniques. In this case, the observation could be enriched by prediction, and as a result, a more comprehensive understanding of the quality of the machined surface could be gained, and at the same time, the cost of study could be reduced. The experimental data (observed)

is used as the training dataset in the machine learning technique. Supervised machine learning technique with some algorithms is then utilized as a powerful tool to recognize the relationship between process variables (input) and response variables (output). In a study of surface roughness under hard turning, for example, cutting speed (v), feed (f), and depth of cut (a) are the process variables, while surface roughness (R_a) is the response. Such supervised machine learning technique algorithms are successfully applied in machining [14, 16–18].

Considering the review presented above, in the work reported in this paper, an experimental study of surface roughness when hard turning of AISI 4340 steels (hardness number of 50 HRC) under MQL conditions is enriched with a machine learning technique. The novel in this study is that the influence of tool wear as an uncontrollable factor in the turning process is considered via cutting length. Predictive models are developed using 2 (two) machine learning algorithms, i.e., Random Forest (RF) and Gaussian Process Regression (GPR). The performance of the models is then evaluated in terms of the errors in their predictions.

2. Materials and method

AISI 4340 bars of high strength low alloy steel with a hardness number of 50 HRC, dimension diameter of 80 mm, and length of 300 mm are used as the workpiece materials for the turning experiment in this study. The uncoated carbide tool coded DCMT11T304-F2 TP40 is used as the cutting tool, and this insert is rigidly mounted on the tool holder coded SDJCR 1616 H11. The turning experiment uses the MQL technique, and FUCHS Ecocut 1012 ID is used for the MQL cutting fluid. This cutting fluid is a low viscosity cutting oil manufactured with solvent-refined paraffinic base oil.

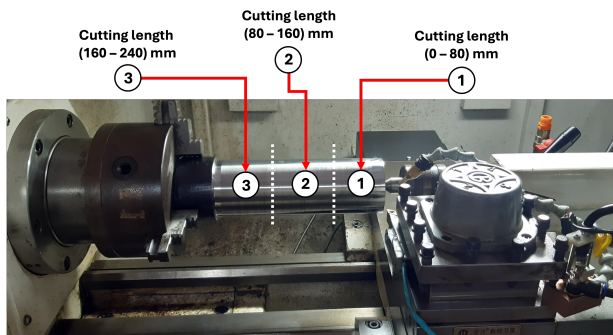
A factorial design is adopted in this experiment, and the factors and the levels are shown in Table 1. Process variables are cutting speed (v), feed (f), depth of cut (a), and as mentioned earlier, tool wear is considered via cutting length (l_t) in this experiment. The preliminary work [19] of this study shows that tool wear of the uncoated carbide tool is associated with cutting length in which the initial wear ($VB < 0.1$ mm), gradual wear ($0.1 < VB < 0.20$ mm), and sudden wear ($VB > 0.2$ mm) phases of flank wear are observed at cutting length of (0 – 80) mm, (80 – 160) mm, and (160–240) mm, respectively.

The response variable in this study is surface roughness. The surface roughness measurement for the R_a parameter is as per ISO standard 1302:2002 [20] and ISO 4288:1998 [21]. Sampling and evaluation length are set accordingly, and a Gaussian filter is used. The instrument used for the

Table 1. Factors and levels of factorial design of the experiment

Factors	Symbols	Units	Levels		
Cutting speed	v	m/min	70	90	110
Feed	f	mm/rev	0.14	0.16	0.22
Depth of cut	a	mm	0.25	-	0.5
Cutting length	lt	mm	0 – 80	80 – 160	160 – 240

observation of machined surface is Dino-Lite AM7115MTF digital microscope and for the measurement of Ra is profilometer Surftest SJ-210. In every cutting condition, as in Table 1, the uncoated carbide tool is assigned for turning with a cutting length of about 240 mm. As the phases of flank wear, the cutting length is divided into 3 (three) phases, with 80 mm for every phase. Surface roughness (Ra) is measured 3 (three) times in every phase, and the average is recorded. Figure 1 shows the arrangement of turning experiment and the position of the three phases of cutting length where the machined surface is observed, and surface roughness (Ra) is measured.

**Fig. 1.** Turning experiment and the position of the three phases of cutting length

There are 2 (two) algorithms of machine learning used in this study they are Random Forest (RF) and Gaussian Process Regression (GPR). The RF algorithm was proposed by Breiman [22] in 2001. This ensemble algorithm improves the predictive accuracy and controls overfitting by averaging multiple decision trees. The advantages of this algorithm include high accuracy, robustness to overfitting, and handling non-linearity well [22, 23].

GPR is a probabilistic supervised machine learning algorithm. It works by extending the concept of the probability distribution of values to functions. GPR calculates the probability distributions over all admissible functions that fit the data. As RF, GPR also works well with a moderate number of experimental runs as a factorial design in this study. RF and GPR are good choices because of their robustness and ability to handle non-linear relationships effectively [24, 25]. The performance of RF and GPR models in this study is

evaluated regarding root mean square error (RMSE) metric- k -fold cross-validation number and size of the training dataset. The RMSE is the normalized distance between the vector of predicted values and the vector of observed values. This metric is mathematically given by:

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

where n is the sample size, $\hat{y}_i$ is a predicted value, and y_i is the corresponding observed value. The smaller the value of RMSE, the higher the predictor's accuracy. The K-fold crossvalidation is a measurement used to assess the performance of a predictive model. In this technique, the sample is divided randomly into K sub-samples. A single sub-sample is regarded as the validation data for testing the model, and the remaining $K - 1$ sub-samples are used as training data. The cross-validation process is repeated K times, and each K sub-sample uses exactly one as the validation data. The K results from the folds can then be averaged to produce a single estimation [26–28].

3. Results and discussion

The result of the hard-turning experiment for Ra measurement on the machined surface at all three phases of cutting length is presented in Table 2. A total of 54 runs were carried out in the experiment.

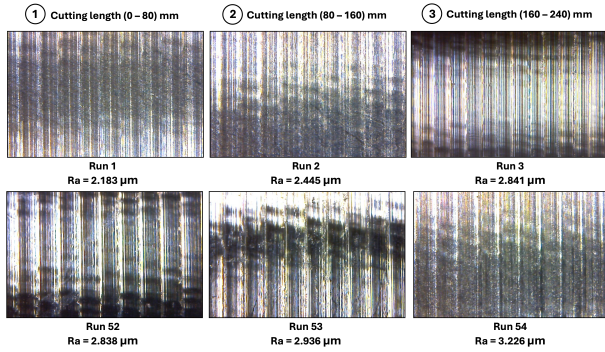
Besides the measurement of surface roughness, the observation was also done on the machined surface at all three phases of cutting length. Some samples of observed machined surface resulted from Runs 1, 2, 3, 52, 53, and 54 are presented in Figure 2. Those observed machined surface were captured using digital microscope with 50x of magnification. From the observation for all 54 machined surface, it is observed that the appearance of machined surface is similar where feed mark (the vertical mark lines in all machined surface photos in Figure 2) is dominantly observed as the texture of machined surface.

3.1 Model training and prediction

Prior to develop the machine learning model, the data in Table 2 was analysed in order to study the relation between the experiment factors (v, f, a, lt) with response variable Ra. The ANOVA analysis reveals that feed (f) and cutting

Table 2. The result of hard turning experiment

v	f	Runs	a	It	$Ra(\mu m)$	Runs	a	It	$Ra(\mu m)$
70	0.14	1	0.25	1	2.183	28	0.5	1	2.413
70	0.14	2	0.25	2	2.445	29	0.5	2	2.745
70	0.14	3	0.25	3	2.841	30	0.5	3	3.165
90	0.14	4	0.25	1	2.192	31	0.5	1	2.598
90	0.14	5	0.25	2	2.395	32	0.5	2	2.725
90	0.14	6	0.25	3	2.835	33	0.5	3	2.978
110	0.14	7	0.25	1	2.325	34	0.5	1	2.485
110	0.14	8	0.25	2	2.623	35	0.5	2	2.714
110	0.14	9	0.25	3	2.933	36	0.5	3	3.322
70	0.16	10	0.25	1	2.425	37	0.5	1	2.42
70	0.16	11	0.25	2	2.723	38	0.5	2	2.867
70	0.16	12	0.25	3	3.282	39	0.5	3	3.244
90	0.16	13	0.25	1	2.355	40	0.5	1	2.285
90	0.16	14	0.25	2	2.825	41	0.5	2	2.795
90	0.16	15	0.25	3	3.334	42	0.5	3	3.256
110	0.16	16	0.25	1	2.225	43	0.5	1	2.788
110	0.16	17	0.25	2	2.765	44	0.5	2	2.865
110	0.16	18	0.25	3	3.212	45	0.5	3	2.935
70	0.22	19	0.25	1	2.625	46	0.5	1	2.644
70	0.22	20	0.25	2	2.978	47	0.5	2	2.973
70	0.22	21	0.25	3	3.365	48	0.5	3	3.385
90	0.22	22	0.25	1	2.434	49	0.5	1	2.384
90	0.22	23	0.25	2	2.833	50	0.5	2	2.637
90	0.22	24	0.25	3	3.178	51	0.5	3	3.328
110	0.22	25	0.25	1	2.688	52	0.5	1	2.838
110	0.22	26	0.25	2	2.751	53	0.5	2	2.936
110	0.22	27	0.25	3	3.482	54	0.5	3	3.226

**Fig. 2.** Samples of machined surface (50x magnification) resulted from Runs 1, 2, 3, 52, 53, and 54

length (l_t) significantly affect surface roughness (Ra), with cutting length contributing the largest percentage of the F-value at 74.6%, highlighting its strong impact. Feed and depth of cut (a) contribute 10.0% and 4.2%, respectively. Cutting speed (v) also influences Ra , with an F-value contribution of 1.5%, indicating a smaller but notable effect on surface roughness. Interactions between factors contribute minimally, suggesting limited combined effects on Ra . By this analysis, it is proof that cutting length

is effectively used as a proxy for tool wear, contributing significantly or dominantly influence to surface roughness prediction.

Data in Table 2 is used as the input for developing machine learning models using RF and GPR algorithms. There are 2 (two) programs have been developed using Python codes for RF and GPR. As the training dataset, the program randomly picks 80% of the input data for model development. The remaining 20% of the input data would be used as the testing dataset to validate the model. The following paragraphs present the results of the model (Ra prediction) and validation (experimental).

The plot of RF and GPR models and validation is presented in Figure 3. The RF model used the training dataset (80% of input data randomly picked from Table 2). As a testing dataset, the remaining 20% of input data (11 runs) was inputted into the RF model to provide the value of Ra prediction. The value of the Ra prediction for 11 runs (7, 8, 21, 27, 33, 40, 42, 49, 50, 51, 53) and the value of the Ra experiment (at the same runs) are plotted in Figure 3(a). For the same purpose of RF but different cutting condition numbers, because runs are randomly picked by the program (2, 3, 5, 8, 13, 15, 23, 25, 35, 44, 48), the plot of the GPR model and its validation are presented in Figure 3(b).

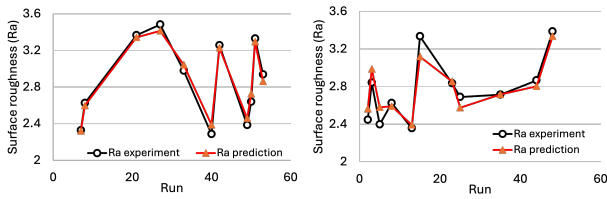


Fig. 3. Plot of model (Ra prediction) and validation (Ra experiment): (a) RF model, and (b) GPR model

The trend of the plot of the Ra experiment and Ra prediction in Figure 3 shows a good agreement. In particular, the plot resulting in the RF model in Figure 3(a), the plot of the Ra experiment and Ra prediction is almost identical. As shown in Table 3, the relative error between the Ra experiment and Ra prediction for the RF model (Figure 3(a)) ranges between (0.23 – 4.45) %, while for the GPR model (Figure 3(b)) is (0.07 – 7.68) %. From this fact, it can be concluded that the prediction provided by the RF model is better than the GPR model.

3.2 Model performance evaluation

As mentioned previously, the performance of models is evaluated in terms of RMSE, Kfold cross-validation number, and size of the training dataset. From the results of model development for RF and GPR algorithms using 5-fold cross-validation and 80% of the training dataset, the RMSE for the RF model is recorded at $0.1746\mu\text{m}$, and the GPR model is $0.2149\mu\text{m}$. This is the reason for the better agreement of the plot (Figure 3(a)) and the smaller relative error provided by the RF model compared to the GPR model, as discussed before. Moreover, the performance evaluation for both models was done by studying the influence of the K -fold number used for cross-validation and the training dataset size. The values of RMSE visualized in Figure 4 are the results of various scenarios arranged to study the influence of the K-fold value and size of the training dataset on the performance of both models.

In Figure 4(a), plots of the RF model at K-5 and K-10 show that the RMSE values vary and are almost identical when the training dataset size is 80% and 90%. At 80%, the RMSE is recorded at the lowest value of RF K-5 (RMSE = $0.1746\mu\text{m}$) and RFK – 10 (RMSE = $0.1732\mu\text{m}$). The case is different for the GPR model, where the RMSE tends to increase with the size of the training dataset increases. Moreover, it must be noted that both GPR K-5 and GPR K-10 have identical RMSE values at all sizes of training datasets. In this case, it seems that K-fold does not influence the RMSE value of the GPR model.

In Figure 4(b), plots of the RF models at the train-

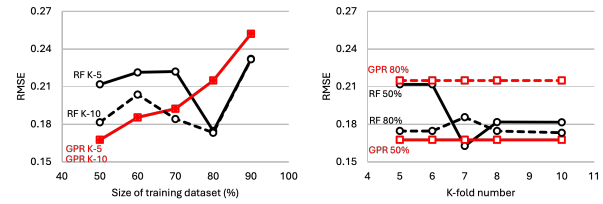


Fig. 4. The RMSE of the RF and GPR models: (a) size of the training dataset (50-90) % at K-fold (5 and 10), and (b) K-fold (5-10) at the size of the training dataset (50 and 80) %

ing dataset of 50% and 80% show that the RMSE values vary, and the lowest value is given by RF 50% at K – 7 (RMSE = $0.1627\mu\text{m}$). In both RF 50% and RF 80%, the RMSE values are constant when K-5 and K-6. Moreover, when K-7, the RMSE value of RF 50% decreases (the lowest value of RMSE = $0.1627\mu\text{m}$), but the RF 80% increases to RMSE = $0.1856\mu\text{m}$. In the remaining, both RF 50% and RF 80% show almost constant values of RMSE at K-8 and K-10. For the GPR model, both GPR 50% and GPR 80% show different but constant values of the RMSE for all K-fold values. This behavior seems to reconfirm the previous result in Figure 4(a) that K-fold is not influencing the RMSE value for the GPR model.

Referring to the works reported in [18, 29–31], the K-fold numbers influenced the result of the error calculation of the model. Motta et al. [18], in modelling of surface roughness when turning of tool steel X37CRMOV5-1, reported that the RMSE values of the GPR model varied between (0.3 to 0.9) with the increase of K number (2 to 10) and size of the training dataset from 50% to 90%. A similar trend was also reported in [29]. Also, for the GPR model, K-5 cross-validation influenced the calculation of RMSE, as reported in [30]. Considering the works of [18, 29, 30], the theory of cross-validation [28], and machine learning programming [25], further analysis was done on the results visualized in Figures 4(a) and 4(b).

The Python code for the GPR model was scrutinized to reconsider the data variability to avoid overfitting or underfitting the GPR model, the random state, the grid parameter, the kernel parameters, and cross-validation to ensure that folds were split as expected. Modifications were made to reduce the grid parameter, enable verbose output to monitor the fitting process and check the diagnostic output. So far, the results showed that the K-fold number influenced the RMSE values at every random state; however, when the RMSE of all random states were averaged, the RMSE values were constant, as shown in Figure 4(b).

The RMSE values from combining every K-fold value

Table 3. Relative error between Ra experiment and Ra prediction for the RF and GPR models

RF model				GPR model			
Runs	Ra(μm)		Relative error (%)	Runs	Ra(μm)		Relative error (%)
	experiment	predicted			experiment	predicted	
7	2.325	2.320	0.23%	2	2.445	2.562	4.79%
8	2.623	2.596	1.03%	3	2.841	2.987	5.14%
21	3.365	3.343	0.66%	5	2.395	2.579	7.68%
27	3.482	3.413	1.99%	8	2.623	2.590	1.27%
33	2.978	3.042	2.15%	13	2.355	2.392	1.56%
40	2.285	2.387	4.45%	15	3.334	3.118	6.49%
42	3.256	3.227	0.88%	23	2.833	2.853	0.71%
49	2.384	2.458	3.09%	25	2.688	2.574	4.25%
50	2.637	2.726	3.37%	35	2.714	2.716	0.07%
51	3.328	3.294	1.02%	44	2.865	2.802	2.21%
53	2.936	2.861	2.57%	48	3.385	3.335	1.48%

with every size of the training dataset were found to be identical. Based on the results, we assume that the GPR model performance is remarkably consistent across different cross-validation splits and random states. Our experimental data is homogeneous; hence, the GPR algorithm's model is very stable. The data's standard deviation and feature importance were also checked to ensure this consistency.

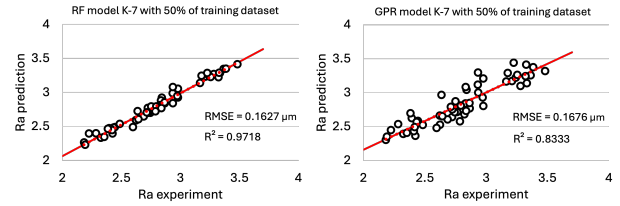
In checking the standard deviation, the results showed that the standard deviation of RMSE for each K-fold value was very low and almost identical for all K-fold values. The difference was recorded after the 10th significant number. This low value indicates that the RMSE values are highly consistent across different random states.

In the case of the feature importances, it was found that the feature cutting length, especially the effect of the cutting length of (180 – 240)mm or *lt3* (in Table 2) to surface roughness (Ra) for all runs, was overwhelmingly the other features in the dataset, contributing 69.5% to the model. The other 5 (five) features (in Table 2) were relatively low importance (*v* = 7.8%, *f* = 7.8%, *a* = 4.6%, *lt1* = 6.3%, and *lt2* = 3.8%) and far less influential than the feature of cutting length of (180-240) mm.

Based on the discussion of the results in the last four paragraphs, the GPR model is stable and not significantly affected by the choice of training and validation splits. Moreover, the GPR model is robust and performs consistently across different cross-validation splits and random states, making it reliable.

The plots in Figure 4(b) show that the lowest RMSE values for both models are recorded at K-7 with 50% training dataset where the RMSE = 0.1627 μm provided by the RF model, and the RMSE = 0.1676 μm provided by the GPR model. Since a smaller RMSE value provides higher predictor accuracy, these results are selected to represent the

models' performance for predicting the surface roughness value (Ra prediction). For visualization of the model's performance, the scatter plots of the Ra experiment versus the corresponding Ra prediction resulting from the models are presented in Figure 5.

**Fig. 5.** Scatter plot of Ra experiment vs. the corresponding Ra prediction resulted by: (a) RF model, and (b) GPR model at K-7 with 50% training dataset

Through the error analysis between the Ra experiment and Ra prediction shown in Figure 5, the relative error of the RF model (Figure 5(a)) is obtained, ranging between (0.17-7.52) %, while for the GPR model (Figure 5(b)) is (0.06-12.93) %. The relative error provided by the RF model is lower than the GPR model. As a reference, a line corresponding to the equality between Ra experiment and Ra prediction shows that the average R^2 score obtained for the RF model ($R^2 = 0.9718$) is higher than the GPR model ($R^2 = 0.8333$). Known as the coefficient of determination, the R^2 score indicates the percentage of the response variable variation that a regression model explains. A higher score of R^2 is preferred to show that a better regression model fits the data. From this result, the conclusion can be given that the performance of the RF model is better than that of the GPR model in predicting the surface roughness value for the problem in this study. Despite the GPR model's slightly lower performance, it remains a viable option for predicting surface roughness. The better per-

formance of RF model was also reported in modelling the surface roughness when turning of AISI 304 steel under MQL technique [32]. Instead of GPR, it was reported in [32] that RF outperformed the other two models (Linear Regression and Support Vector Machine). For 27 data in which 2/3 randomly was used as training dataset and 1/3 for testing, the performance of RF in modelling surface roughness resulted under MQL technique with two types of MQL fluids was indicated by RMSE of $0.1493\mu\text{m}$ and R^2 of 0.9710 for fluid with particle of 30 nm, and RMSE of $0.1396\mu\text{m}$ and R^2 of 0.7968 for fluid with particle of 40 nm.

4. Conclusions

The objective of this study was to explore predictive modeling of surface roughness in the hard turning of AISI 4340 steel using Random Forest (RF) and Gaussian Process Regression (GPR) models. The experimental design considered varying cutting speeds, feeds, depths of cut, and cutting lengths to account for tool wear as an uncontrollable factor. Our findings reveal that cutting length is effectively used as a proxy for tool wear, contributing significantly or dominantly influence to surface roughness prediction, and both models effectively predict surface roughness, with the RF model achieving an RMSE of $0.1627\mu\text{m}$ and an R^2 value of 0.9718, outperforming the GPR model, which had an RMSE of $0.1676\mu\text{m}$ and an R^2 value of 0.8333.

The novelty of this study lies in the inclusion of cutting length as a feature to account for tool wear, significantly impacting the RMSE of the GPR model and underscoring its overwhelming importance compared to other features. Furthermore, using K-fold crossvalidation ($K = 7$) on a 50% training dataset resulted in the lowest RMSE values for both RF and GPR models, highlighting the crucial role of training dataset size in model performance. The GPR model demonstrated robustness and consistency across different cross-validation splits and random states, making it a reliable option despite its slightly lower performance than the RF model. These results suggest that the RF model is more suitable for predicting surface roughness in hard turning operations under the conditions tested. However, the GPR model remains a viable option for scenarios where understanding the underlying distribution of the data is crucial.

The study also identifies the constant RMSE value problem in the GPR model, which is likely related to implementing K-fold cross-validation due to low standard deviation and the significant influence of the cutting length feature. These findings have practical implications for the machining industry, offering a more accurate method for predicting surface roughness, leading to improved product

quality. However, the study has limitations, including its focus on AISI 4340 steel and specific cutting conditions. Future research should validate these models across different materials and a broader range of cutting parameters to ensure their robustness and versatility. Additionally, real-time data acquisition and adaptive learning techniques could enhance predictive capabilities.

Acknowledgments

The research reported in this paper is partially supported by Universitas Sumatera Utara through Grant Number 4167/UN5.1.R/PPM. Moreover, the international research collaboration with Universiti Sultan Zainal Abidin (Terengganu, Malaysia) and Universiti Putra Malaysia (Selangor, Malaysia), is acknowledged.

References

- [1] W. König, R. Komanduri, H. Toenshoff, and G. Ackershott, (1984) "Machining of hard materials" **CIRP annals** 33(2): 417–427. DOI: [10.1016/s0007-8506\(16\)30164-0](https://doi.org/10.1016/s0007-8506(16)30164-0).
- [2] H. K. Toenshoff, C. Arendt, and R. B. Amor, (2000) "Cutting of hardened steel" **Cirp Annals** 49(2): 547–566. DOI: [10.1016/S0007-8506\(07\)63455-6](https://doi.org/10.1016/S0007-8506(07)63455-6).
- [3] R. de Souza Ruzzi, L. R. R. da Silva, R. B. da Silva, W. M. da Silva Junior, and E. C. Bianchi, (2020) "Topographical analysis of machined surfaces after grinding with different cooling-lubrication techniques" **Tribology International** 141: 105962. DOI: [10.1016/j.triboint.2019.105962](https://doi.org/10.1016/j.triboint.2019.105962).
- [4] K. S. S. Lopes, W. F. Sales, and E. S. Palma, (2008) "Influence of machining parameters on fatigue endurance limit of AISI 4140 steel" **Journal of the Brazilian Society of Mechanical Sciences and Engineering** 30: 77–83. DOI: [10.1590/S1678-58782008000100011](https://doi.org/10.1590/S1678-58782008000100011).
- [5] Y. L. Lee, J. Pan, R. B. Hathaway, and M. E. Barkey. *Fatigue testing and analysis: Theory and practice*. 2004. DOI: [10.1016/B978-0-7506-7719-6.X5000-3](https://doi.org/10.1016/B978-0-7506-7719-6.X5000-3).
- [6] W. F. Sales, J. Schoop, L. R. da Silva, Á. R. Machado, and I. S. Jawahir. *A review of surface integrity in machining of hardened steels*. 2020. DOI: [10.1016/j.jmapro.2020.07.040](https://doi.org/10.1016/j.jmapro.2020.07.040).
- [7] S. S. Sarjana, I. Bencheikh, M. Nouari, and A. Ginting, (2020) "Study on cutting performance of cermet tool in turning of hardened alloy steel" **International Journal of Refractory Metals and Hard Materials** 91: DOI: [10.1016/j.ijrmhm.2020.105255](https://doi.org/10.1016/j.ijrmhm.2020.105255).

- [8] M. H. S. Elmunafi, D. Kurniawan, and M. Noordin, (2015) "Use of Castor Oil as Cutting Fluid in Machining of Hardened Stainless Steel with Minimum Quantity of Lubricant" **Procedia CIRP** 26: 408–411. DOI: [10.1016/j.procir.2015.03.001](https://doi.org/10.1016/j.procir.2015.03.001).
- [9] A. E. Diniz, J. R. Ferreira, and F. T. Filho, (2003) "Influence of refrigeration/lubrication condition on SAE 52100 hardened steel turning at several cutting speeds" **International Journal of Machine Tools and Manufacture** 43: DOI: [10.1016/S0890-6955\(02\)00186-4](https://doi.org/10.1016/S0890-6955(02)00186-4).
- [10] D. A. Axinte and R. C. Dewes, (2002) "Surface integrity of hot work tool steel after high speed milling-experimental data and empirical models" **Journal of Materials Processing Technology** 127: DOI: [10.1016/S0924-0136\(02\)00282-0](https://doi.org/10.1016/S0924-0136(02)00282-0).
- [11] Y. B. Guo, W. Li, and I. S. Jawahir, (2009) "Surface integrity characterization and prediction in machining of hardened and difficult-to-machine alloys: A state-of-art research review and analysis" **Machining Science and Technology** 13: DOI: [10.1080/10910340903454922](https://doi.org/10.1080/10910340903454922).
- [12] F. Aggogeri, N. Pellegrini, and F. L. Tagliani. *Recent advances on machine learning applications in machining processes*. 2021. DOI: [10.3390/app11188764](https://doi.org/10.3390/app11188764).
- [13] T. Segreto and R. Teti, (2019) "Machine learning for in-process end-point detection in robot-assisted polishing using multiple sensor monitoring" **International Journal of Advanced Manufacturing Technology** 103: DOI: [10.1007/s00170-019-03851-7](https://doi.org/10.1007/s00170-019-03851-7).
- [14] J. V. Abellan-Nebot and F. R. Subirón. *A review of machining monitoring systems based on artificial intelligence process models*. 2010. DOI: [10.1007/s00170-009-2191-8](https://doi.org/10.1007/s00170-009-2191-8).
- [15] V. Nasir and F. Sassani. *A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges*. 2021. DOI: [10.1007/s00170-021-07325-7](https://doi.org/10.1007/s00170-021-07325-7).
- [16] B. hu Li, B. cun Hou, W. tao Yu, X. bing Lu, and C. wei Yang. *Applications of artificial intelligence in intelligent manufacturing: a review*. 2017. DOI: [10.1631/FITEE.1601885](https://doi.org/10.1631/FITEE.1601885).
- [17] J. F. Arinez, Q. Chang, R. X. Gao, C. Xu, and J. Zhang, (2020) "Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook" **Journal of Manufacturing Science and Engineering, Transactions of the ASME** 142: DOI: [10.1115/1.4047855](https://doi.org/10.1115/1.4047855).
- [18] M. P. Motta, C. Pelaingre, A. Delamézière, L. B. Ayed, and C. Barlier. "Machine learning models for surface roughness monitoring in machining operations". In: *Procedia CIRP*. 108. 2022. DOI: [10.1016/j.procir.2022.03.110](https://doi.org/10.1016/j.procir.2022.03.110).
- [19] R. A. Sidabutar. "Pengembangan sistem dan unit Minimum Quantity Lubrication (MQL) untuk meningkatkan performa pahat pada pemesinan keras". (mathesis). Universitas Sumatera Utara, 2021.
- [20] I. 1302:2002. ISO 1302-2002: Geometrical Product Specifications (GPS) - Indication of surface texture in technical product documentation. Tech. rep. ISO International Standard, 2002.
- [21] I. 4288:1998. Geometrical Product Specifications (GPS) — Surface texture: Profile method — Rules and procedures for the assessment of surface texture. Tech. rep. ISO International Standard, 1998.
- [22] L. Breiman, (2001) "Random forests" **Machine Learning** 45: 5–32.
- [23] M. Schonlau and R. Y. Zou, (2020) "The random forest algorithm for statistical learning" **The Stata Journal: Promoting communications on statistics and Stata** 20: 3–29. DOI: [10.1177/1536867X20909688](https://doi.org/10.1177/1536867X20909688).
- [24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, (2011) "Scikit-learn: Machine learning in Python" **Journal of Machine Learning Research** 12:
- [25] A. C. Müller and S. Guido. *Introduction to Machine Learning with Python and Scikit-Learn*. 2015.
- [26] P. Jiang and J. Chen, (2016) "Displacement prediction of landslide based on generalized regression neural networks with K-fold cross-validation" **Neurocomputing** 198: DOI: [10.1016/j.neucom.2015.08.118](https://doi.org/10.1016/j.neucom.2015.08.118).
- [27] A. Marrel and B. Iooss, (2024) "Probabilistic surrogate modeling by Gaussian process: A review on recent insights in estimation and validation" **Reliability Engineering & System Safety** 247: 110094. DOI: [10.1016/j.res.2024.110094](https://doi.org/10.1016/j.res.2024.110094).
- [28] T. Hastie, R. Tibshirani, and J. Friedman, (2009) "The Elements of Statistical Learning, Second Edition" Springer New York, NY 27:

- [29] B. Anandan and M. Manikandan, (2023) “Machine learning approach with various regression models for predicting the ultimate tensile strength of the friction stir welded AA 2050-T8 joints by the K-Fold cross-validation method” **Materials Today Communications** 34: 105286. DOI: [10.1016/j.mtcomm.2022.105286](https://doi.org/10.1016/j.mtcomm.2022.105286).
- [30] A. Rohani, M. Taki, and M. Abdollahpour, (2018) “A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I)” **Renewable Energy** 115: 411–422. DOI: [10.1016/j.renene.2017.08.061](https://doi.org/10.1016/j.renene.2017.08.061).
- [31] G. Jiang and W. Wang, (2017) “Error estimation based on variance analysis of k -fold cross-validation” **Pattern Recognition** 69: 94–106. DOI: [10.1016/j.patcog.2017.03.025](https://doi.org/10.1016/j.patcog.2017.03.025).
- [32] V. Dubey, A. K. Sharma, and D. Y. Pimenov, (2022) “Prediction of Surface Roughness Using Machine Learning Approach in MQL Turning of AISI 304 Steel by Varying Nanoparticle Size in the Cutting Fluid” **Lubricants** 10: DOI: [10.3390/lubricants10050081](https://doi.org/10.3390/lubricants10050081).