SEMARAK ILMU PUBLISHING

Semarak Engineering Journal

Journal homepage:
https://semarakilmu.my/index.php/sej/index
ISSN: 3036-0145



Performance of AISI 1045 Axial Surface Roughness Classification Using K-Nearest Neighbor with Multiple Distance Metrics and an Ensemble Method

M. A. Ade Saputra[1], Irsyadi Yani[2,*], Yulia Resti[3], Dewi Puspitasari[2], Ismail Thamrin[2], Muhammad Yanis[2]

[1] Faculty of Engineering, Universitas Sriwijaya, Palembang, Sumatera Selatan, Indonesia
[2] Department of Mechanical Engineering, Faculty of Engineering, Universitas Sriwijaya, Indralaya 30662, South Sumatra, Indonesia
[3] Department of Mathematic, Faculty of Math and Natural Science, Universitas Sriwijaya, Indralaya 30662, South Sumatra, Indonesia

ARTICLE INFO

Article history:
Received 2 September 2026
Received in revised form 30 October 2026
Accepted 1 November 2026
Available online 18 November2026

Keywords:
AISI 1045; distance metric; ensemble learning; k-nearest neighbors; surface roughness classification.

ABSTRACT

Reliable classification of surface-roughness grades is important for machining quality control because incorrect decisions may increase rework, inspection time, and production cost. This study evaluates k-nearest-neighbor (KNN) classifiers for predicting the axial surface-roughness grades of face-milled AISI 1045 steel using Euclidean, Cosine, and Hamming distance metrics. A hard-voting ensemble combining the three classifiers is also examined. The predictor set comprises cutting speed, feed rate, depth of cut, and tangential roughness measured at six locations. Candidate neighborhood sizes were evaluated from k = 1 to 20, and the selected models were subsequently assessed through repeated hold-out resampling with replacement. Model performance was quantified using Accuracy, Precision, Recall, F1-score, and area under the receiver-operating-characteristic curve (AUC), followed by one-way analysis of variance and Tukey-Cramer comparisons. Euclidean KNN was the strongest and most stable individual classifier, whereas the ensemble achieved the highest mean Accuracy (83.01%), Recall (59.45%), F1-score (55.72%), and AUC (73.33%). Euclidean KNN retained the highest Precision (55.02%). The results demonstrate that combining complementary neighborhood geometries improves detection-oriented performance, although class imbalance and the limited representation of the N3 roughness grade remain important constraints. Relative to Euclidean KNN, the ensemble improved Accuracy by only 1.52 percentage points and F1-score by 1.51 points while reducing Precision by 1.34 points. These results reflect internal resampling under a single machining condition; computational cost, external validation, and real-time feasibility were not established.

* *Corresponding author.*
E-mail address: irsyadiyani@ft.unsri.ac.id

https://doi.org/10.37934/sej.17.1.116

1. Introduction

Surface roughness is a critical indicator of machining quality because it affects dimensional conformity, friction, fatigue resistance, sealing performance, and the need for secondary finishing operations [1]. For AISI 1045 steel, which is widely used for shafts, gears, pins, and other load-bearing components, reliable roughness assessment can support process control and reduce avoidable inspection and rework. Machine-learning methods have therefore become increasingly relevant for mapping machining conditions and measured surface features to discrete quality grades [2]–[5].

K-nearest neighbors (KNN) is a nonparametric supervised-learning method that assigns a test observation to the class most frequently represented among its k closest training observations [6]–[8]. Its behavior is governed by two closely related decisions: the number of neighbors and the metric used to define proximity. Euclidean distance emphasizes absolute differences between normalized continuous variables, Cosine distance emphasizes vector orientation, and Hamming distance counts mismatched coordinates after a suitable discrete representation [9]–[13]. Consequently, identical observations may receive different neighborhood structures and class labels under different metrics.

Previous investigations have used decision trees and other machine-learning methods to predict machining surface roughness [3], [4], [14]–[16]. In particular, Resti et al. [14] examined axial roughness grades in wet-milled AISI 1045 using crisp/fuzzy discretization and decision-tree voting. The present work does not propose a novel KNN algorithm or voting rule. Instead, the narrower research question is whether changing neighborhood geometry (continuous-magnitude, vector-orientation, or binary mismatch) changes grade-detection performance and whether conventional hard voting offers a practically worthwhile increment over a single Euclidean KNN. This is a comparative, application-specific methodological assessment rather than an algorithmic advance.

Accordingly, the study (1) compares Euclidean, Cosine, and binary-encoded Hamming KNN using the stated machining and tangential-roughness predictors; (2) evaluates an unweighted three-model hard vote as a benchmark rather than a new ensemble architecture; (3) describes variation across 50 repeated 80:20 hold-out evaluations; and (4) examines Accuracy alongside class-sensitive Precision, Recall, F1-score, and AUC. Its scientific contribution lies in documenting the performance trade-offs and limits of these established choices on axial AISI 1045 roughness grading. Practical utility is evaluated cautiously because predictive gains alone do not establish inspection efficiency or production readiness.

2. Related Works

Distance selection is central to KNN because the nearest-neighbor set changes with the geometry imposed on the predictor space. Manhattan distance accumulates absolute coordinate differences and can be useful when deviations along individual dimensions should contribute linearly. Nevertheless, it may overestimate practical separation in some high-dimensional or manifold-structured data [14]. Euclidean distance measures straight-line separation and remains the most common choice for continuous-valued predictors [13], [17], [18]. Its performance, however, depends strongly on feature scaling because variables with larger numerical ranges can dominate the distance calculation [8].

Cosine distance compares the orientation of two predictor vectors rather than their absolute magnitude. This property can be advantageous when relative patterns are more informative than overall level [19]. Hamming distance counts coordinate mismatches and is naturally suited to categorical or binary representations [11], [12], [20]. Its application to continuous machining variables therefore requires discretization or binary conversion, which introduces an additional

modeling decision. The resulting classifier reflects both the distance rule and the information retained during conversion.

The neighborhood size k controls the bias-variance trade-off. Small values can closely follow local structure but may be sensitive to noise, outliers, and isolated observations. Large values smooth local decision boundaries and may favor majority classes by incorporating distant neighbors [21]–[23]. Because the optimal k is dataset- and metric-dependent, it should be selected empirically using a validation procedure that is separate from the final performance assessment.

Ensemble learning combines predictions from multiple base classifiers through voting or averaging [24]. Performance gains are most likely when the component models are individually informative and make partially different errors. A voting ensemble does not automatically outperform its strongest member; its value depends on error diversity and on whether the voting rule improves the performance criterion relevant to the application. In the present problem, Euclidean, Cosine, and Hamming KNN provide different representations of neighborhood similarity and are therefore plausible complementary base models.

Machine-learning applications in machining commonly use process settings and measured surface features as predictors [3]–[5], [16], [24]–[26]. The most directly related publication is Resti et al. [14], which predicts axial AISI 1045 roughness grades from wet-milling measurements using individual decision trees with crisp or fuzzy discretization and ensembles of decision trees. By contrast, this study compares nearest-neighbor classifiers based on Euclidean, Cosine, and binary-encoded Hamming distances and combines their labels by hard voting. Thus, the principal methodological distinction is neighborhood-based similarity versus tree-based decision rules, not the invention of a new ensemble principle. Both studies use closely matching material, machining conditions, measured predictors, response type, and overlapping authorship. Whether their underlying observation records are identical, partially overlapping, or independently collected cannot be verified from the information provided in this manuscript; the authors must disclose the precise dataset relationship before submission.

The unresolved question is therefore not whether voting or KNN is new, but whether distance representation yields a consistent and useful accuracy–detection trade-off for this particular machining-quality task. Unlike [14], which emphasizes crisp/fuzzy tree discretization, the present comparison isolates three KNN input-distance configurations and evaluates their aggregate predictions. Numerical comparisons with [14] should be interpreted descriptively unless dataset identity, preprocessing, data partitions, and evaluation definitions are demonstrably matched; they are not evidence of independent cross-study superiority.

3. Methodology

The study integrates face-milling experiments, surface-roughness measurement, predictor preprocessing, three KNN classifiers, and a hard-voting ensemble. Candidate neighborhood sizes from k = 1 to 20 were examined separately for Euclidean, Cosine, and Hamming distance. The model-development stage used the repeated resampling procedure illustrated in Figure 1. After the neighborhood sizes had been fixed, the final comparative evaluation was repeated over 50 random hold-out partitions. This separation reduces reliance on a single favorable split and enables the stability of each model to be quantified.

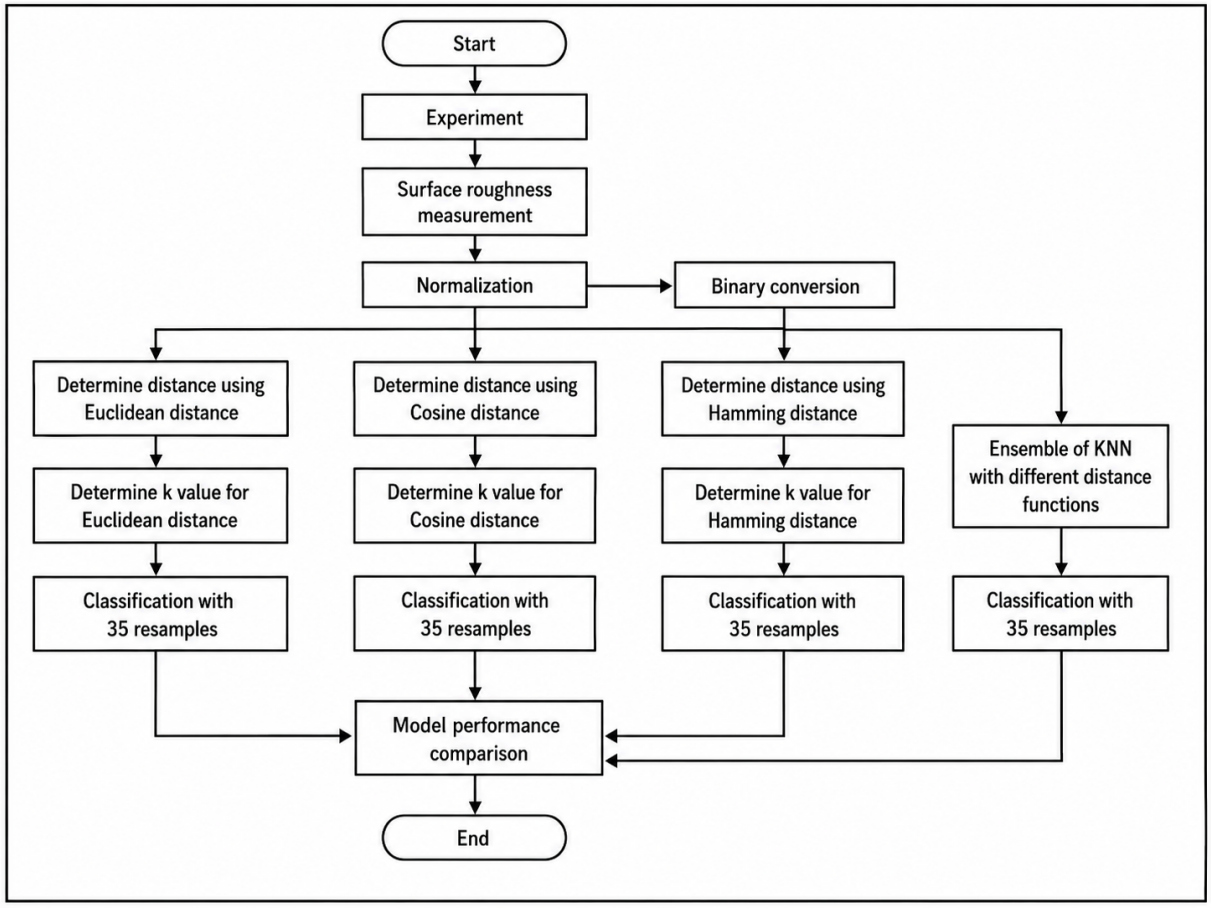


Fig. 1. Flowchart of the research methodology

Face milling was performed in the down-milling direction on AISI 1045 steel specimens measuring 200 × 100 × 25 mm with a hardness of 40-45 HRC. Bromus cutting fluid was used under wet-machining conditions. The cutting tool was a coated carbide end mill with a hardness of 90-93 HRC. The controlled machining variables were cutting speed (V_c), feed rate (f_z), and axial depth of cut (a_x). Axial roughness was measured using a Handysurf Accretech E-35B with a nominal resolution of 0.01 µm, an evaluation length of 4 mm, and a cutoff length of 0.8 mm. Six tangential roughness measurements, R_{t1}-R_{t6}, were also recorded at randomly selected locations and included as predictors. Thus, every observation contained nine input variables and one axial roughness-grade label.

Experimental provenance relative to reference [14]: the materials, wet-machining configuration, specimen dimensions, roughness instrument, and predictor descriptions closely resemble those reported previously. The present submission does not establish whether the underlying observations were newly acquired or reused. Before publication the authors should explicitly state the acquisition dates, number of unique specimens and observations, any shared observation identifiers or percentages of overlap with [14], and whether additional measurements were obtained. Until verified, this work should not be represented as independent experimental validation of [14].

3.1 Data Representation, KNN Models, and Ensemble Rule

Each observation contains nine predictors: V_c, f_z, a_x, and R_{t1}-R_{t6}. Continuous predictors were normalized before the Euclidean and Cosine distances were calculated so that variables with wider numerical ranges did not dominate the neighborhood structure. For Hamming KNN, the normalized predictors were converted to binary form before distance calculation. This conversion allows

Hamming distance to be applied but also reduces within-bin magnitude information; the corresponding results must therefore be interpreted as the combined effect of binary encoding and Hamming distance.

1. Non-negativity: d(A,B) ≥ 0
2. Identity: d(A,B) = 0 if and only if A = B
3. Symmetry: d(A,B)=d(B,A)

Triangle Inequality: d(A,B) + d(B,C) ≥ d(A,B) = d(A,C)

For a test observation, KNN identifies the k training observations with the smallest distance and assigns the roughness grade that occurs most frequently among those neighbors [12], [24], [27]–[29]. Candidate values k = 1, 2, ..., 20 were assessed separately for each distance metric. The selected k value was based on the joint behavior of Accuracy, Precision, Recall, F1-score, and AUC rather than on a single metric. This multi-metric criterion was adopted because maximizing Accuracy alone can favor majority grades and obscure weak minority-class detection.

Since there is no universally accepted rule for selecting the value of k for each distance function, k is typically determined experimentally [22]. In this work, experiments were conducted to assess the performance of surface roughness prediction metrics using three distance functions across k values ranging from 1 to 20. The k value yielding the highest dominant performance metric was selected for each distance function.

Let (x_{1a},x_{2a},⋯,x_{pa}) and B = (x_{1b},x_{2b},⋯,x_{pb}), where m=1, 2, ⋯, p, denote two observations with p predictors. The predictors comprise the three machining variables and six tangential roughness measurements. Euclidean, Cosine, and Hamming distances are defined in (1)-(3), respectively.

Euclidean distance retains normalized magnitude differences, Cosine distance emphasizes the angular relationship between predictor vectors, and Hamming distance counts coordinate mismatches after binary conversion. These different geometries are expected to generate partially different neighborhood memberships and classification errors.

Euclidean distance is appropriate for normalized continuous attributes because all coordinate differences contribute directly to the neighborhood geometry [10], [18], [24]. The normalization step is essential in this study because the tangential roughness measurements and machining variables are expressed on different numerical scales.

$$\mathrm{d(A,B)}=\sqrt{\sum_{m=1}^{p}\left(x_{ma}-x_{mb}\right)^{2}} \tag{1}$$

Cosine distance is obtained from one minus the cosine similarity between two observations [24]. It emphasizes proportional similarity and is less sensitive to the overall magnitude of a predictor vector. The expression used in this study is given in (2).

$$\mathrm{d(A,B)}=1-\frac{\sum_{m=1}^{p}x_{ma}x_{mb}}{\sqrt{\sum_{m=1}^{p}x_{ma}{}^{2}}\sqrt{\sum_{m=1}^{p}x_{mb}{}^{2}}} \tag{2}$$

Hamming distance measures the number of coordinates at which two encoded observations differ [11]. The expression is given in (3). Because the original predictors are continuous, the use of Hamming distance requires binary conversion and may discard small but physically meaningful differences between observations.

$$d(A,B)=\sum_{m=1}^{p} 1_{x_{ma} \neq x_{mb}} \quad (3)$$

The Hamming model should therefore not be interpreted as a metric-only comparison with Euclidean and Cosine KNN. Its performance represents a combined encoding-and-distance configuration. This distinction is relevant when explaining both its different k-response pattern and its comparatively high variability.

The ensemble combines the class labels predicted by Euclidean, Cosine, and Hamming KNN through unweighted hard voting [24]. With three component classifiers, a unique majority label is obtained whenever at least two classifiers agree. The ensemble is intended to improve robustness by retaining the continuous magnitude information of Euclidean KNN, the orientation-based similarity of Cosine KNN, and the discretized mismatch structure of Hamming KNN.

3.2 Performance Metrics and Statistical Evaluation

For each roughness grade, one-versus-rest true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were calculated. Accuracy summarizes overall correctness, Precision measures the proportion of predicted members of a grade that are correct, Recall measures the proportion of actual members detected, and F1-score is the harmonic mean of Precision and Recall. AUC summarizes ranking performance across decision thresholds. Because the task is multiclass and the grade frequencies are unequal, the five metrics were interpreted jointly rather than treating Accuracy as the sole decision criterion.

Statistical interpretation requires caution: the same finite experimental dataset was repartitioned repeatedly, so the 50 evaluations are correlated rather than 50 independent production trials. One-way ANOVA across model-level resampling summaries and a Tukey post-hoc test do not by themselves establish superiority of the ensemble over Euclidean KNN under paired repeated splits. A paired comparison of within-split performance differences, with confidence intervals and a clearly specified class-averaging convention, should be conducted using the original prediction records. Also, scaling, binary thresholds, and k selection should be estimated solely within each training fold to exclude information leakage; the existing description does not establish all these details.

The final comparison used 50 repeated 80:20 hold-out partitions with replacement. For each model and metric, the mean, standard deviation, minimum, and maximum were calculated. One-way ANOVA at $\alpha = 0.05$ was used to test the null hypothesis that the four model means were equal. When the omnibus test was significant, Tukey-Cramer comparisons were used to identify the model pairs responsible for the difference. The statistical analysis was applied separately to Accuracy, Precision, Recall, F1-score, and AUC.

3. Results and Discussion

Table 1 summarizes the predictor ranges used in model development. The six tangential roughness variables span substantially wider and nonuniform ranges compared with the three machining variables. This scale heterogeneity justifies normalization before Euclidean and Cosine distance calculation. It also indicates that the measured surface state contributes information beyond the nominal machining settings, because R_{t1}-R_{t6} capture local variation across the milled surface.

Figure 2 presents the five performance metrics obtained for k values from 1 to 20. The curves confirm that neighborhood size cannot be selected independently of the distance metric. Small

changes in k alter the class composition of the local neighborhood and can produce marked changes in Precision, Recall, and F1-score, particularly when the test set contains few observations from one or more roughness grades.

Table 1
Predictor variable

Predictor Variable	Range
Rt_1	0.4 – 16.7 µm
Rt_2	0.5 – 23.8 µm
Rt_3	0.4 – 16.4 µm
Rt_4	0.33 – 14.72 µm
Rt_5	0.4 – 18.7 µm
Rt_6	0.35 – 32.3 µm
V_c	7.5 – 20 m/min
f_z	0.05 - 0.15 mm/g
a_x	0.75 – 1.5 mm

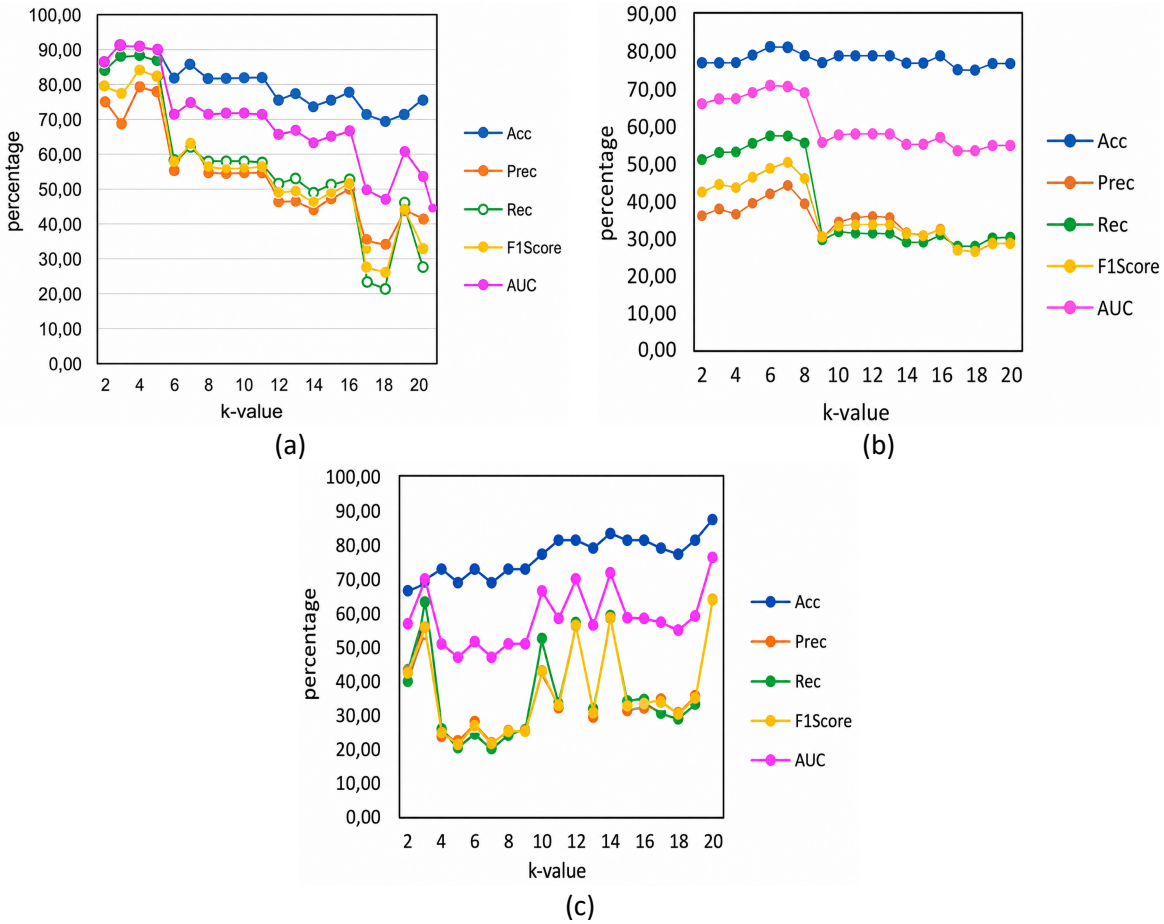


Fig. 2. The k-value for each of the Distance Function: (a) Euclidean, (b) Cosine, and (c) Hamming

The Euclidean model performs best at a relatively small neighborhood size. At k = 4, Precision and F1-score reach their highest values, while Accuracy and Recall are equal to those obtained at k = 3 and AUC is only slightly lower. Therefore, k = 4 provides a reasonable multi-metric compromise. For Cosine KNN, k = 7 was selected because it provides the strongest combined behavior across the reported metrics, although Accuracy and Recall peak at k = 6. Hamming KNN exhibits a distinctly more

irregular response because binary conversion produces abrupt changes in coordinate matching; k = 14 was selected for the subsequent evaluation. The selected values were therefore k = 4, 7, and 14 for Euclidean, Cosine, and Hamming KNN, respectively.

The different k-response patterns are consistent with the geometry of the three metrics. Euclidean distance preserves normalized absolute separation and produces comparatively coherent neighborhoods for continuous machining measurements. Cosine distance groups observations according to proportional patterns, even when their absolute magnitudes differ. Hamming distance operates on the binary representation and can cause several originally distinct observations to become equidistant. This loss of resolution explains why its validation curves fluctuate more strongly and why a larger neighborhood is required to stabilize the prediction.

Figure 3 presents confusion matrices from the first, second, and final resampling iterations for the three selected KNN models. Reporting several iterations is important because the composition of the test subset changes across random partitions. The matrices show that classification errors are concentrated mainly between neighboring roughness grades, especially N5 and N6, rather than between the most widely separated grades. This pattern is technically plausible because adjacent grades are expected to have overlapping machining and tangential-roughness characteristics.

	class	actual			
		N4	N5	N6	N7
Prediction	N4	1	1	0	0
	N5	0	9	3	0
	N6	0	1	8	0
	N7	0	0	0	1

	class	actual			
		N4	N5	N6	N7
Prediction	N4	2	2	0	0
	N5	1	5	4	0
	N6	0	0	9	1
	N7	0	0	0	0

	class	actual			
		N4	N5	N6	N7
Prediction	N4	0	0	0	0
	N5	1	0	3	0
	N6	0	2	4	3
	N7	0	0	1	10

(a)

	class	actual			
		N4	N5	N6	N7
Prediction	N4	1	2	0	0
	N5	0	4	1	0
	N6	0	5	10	1
	N7	0	0	0	0

	class	actual			
		N4	N5	N6	N7
Prediction	N4	0	0	0	0
	N5	2	6	2	1
	N6	1	1	11	0
	N7	0	0	0	0

	class	actual			
		N4	N5	N6	N7
Prediction	N4	0	0	0	0
	N5	1	1	4	0
	N6	0	0	2	4
	N7	0	1	2	9

(b)

	class	actual			
		N4	N5	N6	N7
Prediction	N4	1	0	0	0
	N5	0	8	3	0
	N6	0	3	8	1
	N7	0	0	0	0

	class	actual			
		N4	N5	N6	N7
Prediction	N4	0	0	0	0
	N5	3	3	4	0
	N6	0	4	9	1
	N7	0	0	0	0

	class	actual			
		N4	N5	N6	N7
Prediction	N4	0	0	0	0
	N5	1	1	0	0
	N6	0	1	5	4
	N7	0	0	3	9

(c)

Fig. 3. Confusion matrix of KNN prediction: (a) Euclidean, (b) Cosine, and (c) Hamming

Class N3 is absent from the displayed test predictions because only two N3 observations are available and random partitioning does not guarantee representation in every test subset. This extreme scarcity prevents a reliable class-specific assessment and also explains why macro-averaged Precision, Recall, and F1-score vary more than Accuracy. For the first Euclidean resample, the classwise TP counts for N4-N7 are 1, 9, 8, and 1, respectively. The resulting Accuracy, Precision, Recall, F1-score, and AUC are 89.58%, 78.47%, 88.64%, 83.24%, and 89.93%. These values illustrate why a single split would provide an overly optimistic and unstable estimate; the final conclusions are consequently based on the 50-resample summaries.

Figure 4 shows the metric trajectories over 50 resampling iterations for Euclidean, Cosine, and Hamming KNN. The order of the iterations has no physical meaning; the plots are used to visualize variation across alternative train-test partitions. Table II provides the corresponding means, standard deviations, minima, and maxima and is therefore the primary basis for comparing both average performance and stability.

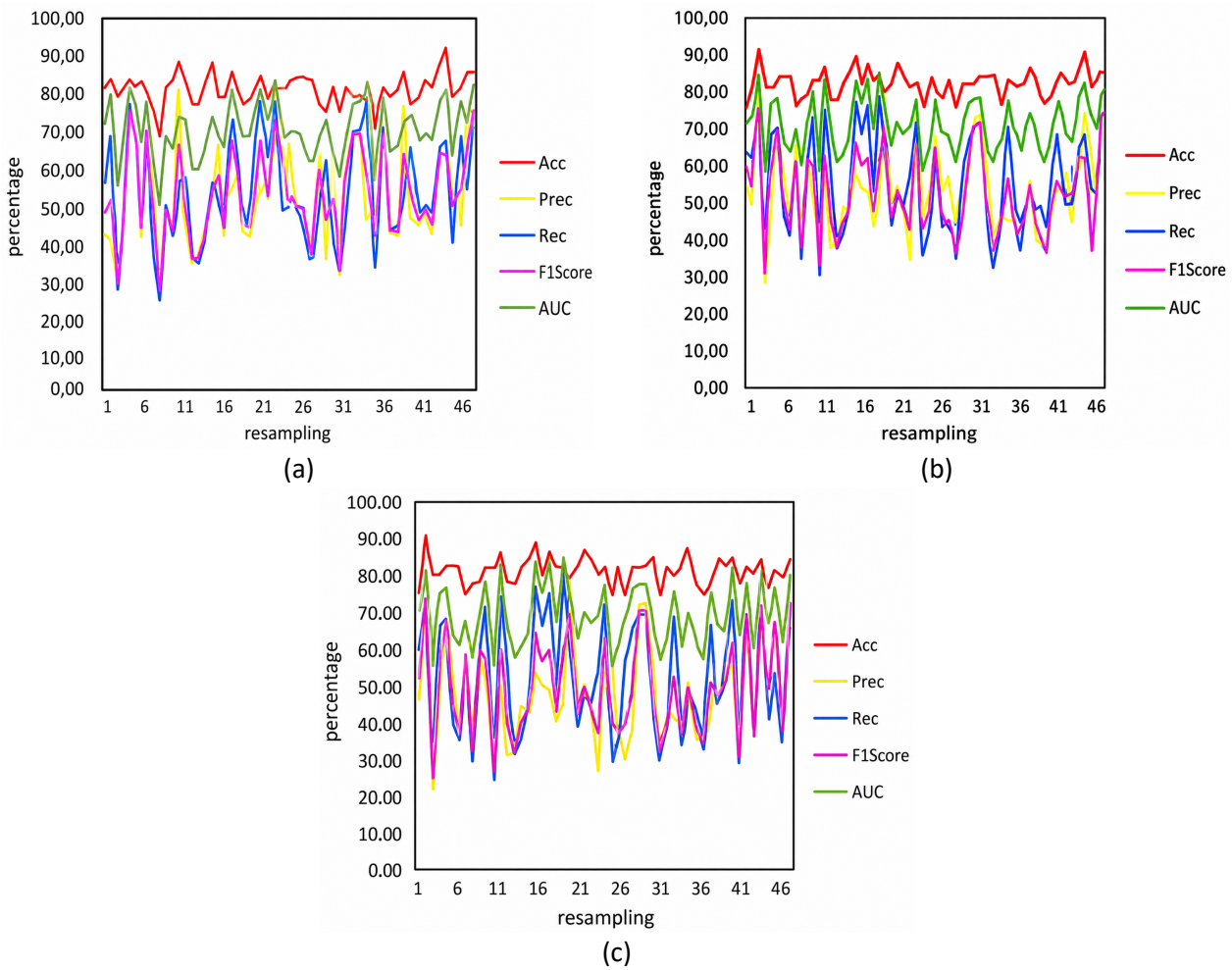


Fig. 4. KNN performance in resampling with replacement: (a) Euclidean, (b) Cosine, and (c) Hamming

For Euclidean KNN, Accuracy has the smallest standard deviation (3.82 percentage points) and range (20.84 points), whereas Precision, Recall, and F1-score vary more strongly. The corresponding coefficients of variation are approximately 4.7% for Accuracy, 24.5% for Precision, 23.2% for Recall, 20.6% for F1-score, and 10.1% for AUC. These results show that the model is comparatively stable in overall correctness but less stable in its treatment of individual grades.

Cosine KNN produces lower mean values than Euclidean KNN for all five metrics. Its standard deviations are 3.99, 14.63, 11.00, 10.90, and 6.33 percentage points for Accuracy, Precision, Recall, F1-score, and AUC, respectively. The largest relative variability again occurs in Precision, Recall, and F1-score. This behavior indicates that orientation-based similarity alone does not consistently separate the axial roughness grades when absolute differences in machining and tangential roughness remain informative.

Hamming KNN exhibits the widest overall ranges: 22.19 percentage points for Accuracy, 67.43 for Precision, 56.86 for Recall, 59.12 for F1-score, and 35.60 for AUC. Its Precision standard deviation reaches 18.02 points, the largest value among all models. These results support the interpretation that binary conversion makes neighborhood membership highly sensitive to the sampled observations and to values near the conversion threshold.

Across the three base classifiers, Euclidean KNN provides the strongest combination of mean performance and stability. Hamming KNN achieves a higher mean Accuracy than Cosine KNN but lower Recall, F1-score, and AUC. Therefore, ranking the models by Accuracy alone would overstate

the quality of the Hamming classifier. The multi-metric evaluation instead identifies Euclidean KNN as the most reliable individual configuration.

Figure 5 presents the performance trajectories of the hard-voting ensemble over the same 50 resampling iterations. The ensemble achieved mean Accuracy, Precision, Recall, F1-score, and AUC values of 83.01%, 53.68%, 59.45%, 55.72%, and 73.33%, respectively. Its standard deviations were 3.36, 11.30, 14.77, 11.25, and 7.93 points. The relatively low Accuracy variation indicates stable overall classification, while the wider Recall variation reflects the changing representation of minority grades across test subsets.

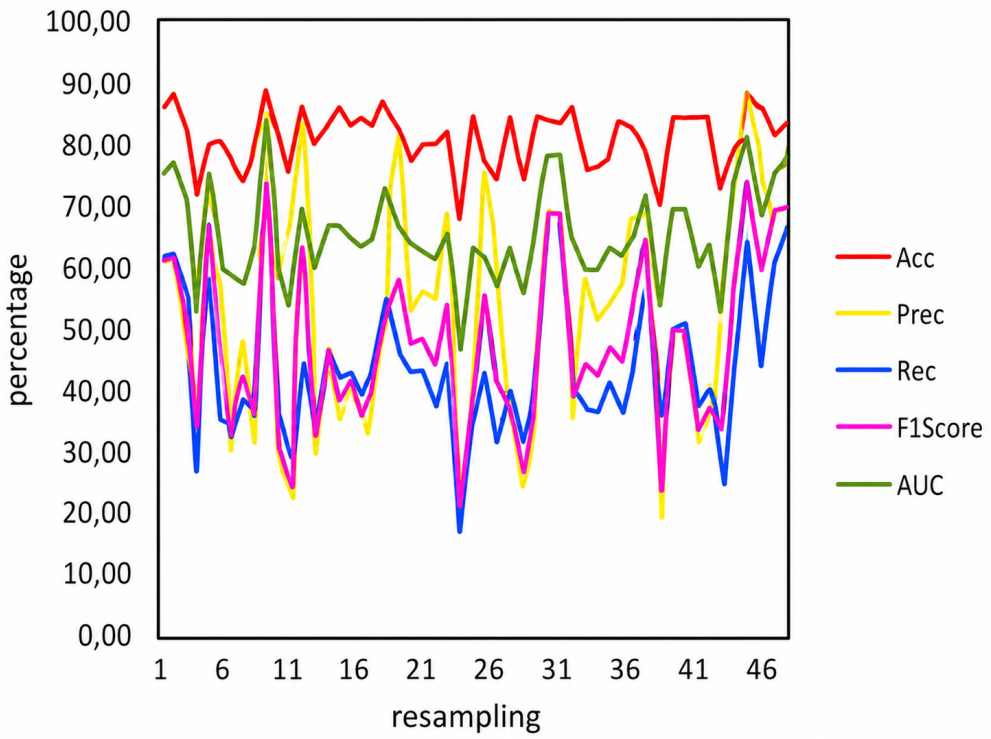


Fig.5. Performance of the ensemble method of KNN in resampling with replacement

Compared with Euclidean KNN, the ensemble raised mean Accuracy from 81.49% to 83.01% (+1.52 percentage points), Recall from 54.88% to 59.45% (+4.57 points), F1-score from 54.21% to 55.72% (+1.51 points), and AUC from 70.65% to 73.33% (+2.68 points); Precision fell from 55.02% to 53.68% (−1.34 points). These are small changes in overall Accuracy and F1-score, and improved Recall may come at the expense of additional false-positive grade assignments. The higher ensemble mean should therefore not be interpreted as proof of a meaningful industrial benefit. The simpler Euclidean model remains a defensible default when model complexity, reproducibility, or Precision are prioritized. The ensemble may merit consideration only if missing an out-of-specification grade has a documented cost high enough to justify its Recall gain.

Complexity–benefit assessment: a Euclidean KNN query ordinarily requires distances to the retained training instances and selection of k neighbors. The proposed ensemble evaluates three separate KNN models (including encoding for Hamming) and performs an additional vote; the exact time and memory overhead depend on the implementation and dataset size. No wall-clock training time, per-part prediction latency, memory utilization, or throughput measurements were collected or reported. Hence the observed +1.52-point Accuracy gain cannot presently be weighed quantitatively against inference cost. A deployment study should benchmark the same hardware and preprocessed input stream, reporting latency distributions, memory demand, parts inspected per

second, and net false-accept/false-reject costs. Moreover, the method requires six measured tangential-roughness inputs; unless these are available within the desired cycle time, the approach is better described as inspection assistance than as demonstrated in-process, real-time prediction.

Compared with Euclidean KNN, the ensemble increased Accuracy by 1.52 percentage points, Recall by 4.57 points, F1-score by 1.51 points, and AUC by 2.68 points, while Precision decreased by 1.34 points. Relative to Cosine KNN, the ensemble improved Accuracy, Recall, F1-score, and AUC by 4.36, 13.25, 7.61, and 8.64 points, respectively. Relative to Hamming KNN, the corresponding gains were 2.73, 15.29, 9.55, and 9.04 points. The ensemble is therefore most beneficial for detection-oriented metrics, but Euclidean KNN remains preferable when Precision and model simplicity are prioritized.

Table 2 confirms that the ensemble has the highest mean Accuracy, Recall, F1-score, and AUC, whereas Euclidean KNN has the highest mean Precision. The model ranking is not uniform across metrics: Euclidean is the strongest base classifier, Cosine and Hamming contribute weaker but different neighborhood structures, and voting improves the balance between correctly detecting roughness grades and controlling overall errors. The reported minimum and maximum values also show that the ensemble does not eliminate partition sensitivity, particularly for Recall.

Table 2
Statistical summary of KNN performance metric

Distance metric	Statistic	Accuracy	Precision	Recall	F1-score	AUC
Euclidean	Mean	81.49	55.02	54.88	54.21	70.65
	Std	3.82	13.48	12.75	11.17	7.15
	Minimum	70.83	32.91	30.00	31.39	56.10
	Maximum	91.67	81.61	80.68	75.60	83.63
Cosine	Mean	78.65	52.09	46.20	48.11	64.69
	Std	3.99	14.63	11.00	10.90	6.33
	Minimum	72.92	25.00	27.75	27.25	53.25
	Maximum	87.50	79.23	80.49	72.98	82.11
Hamming	Mean	80.28	50.42	44.16	46.17	64.29
	Std	4.98	18.02	12.31	13.46	7.57
	Minimum	67.39	20.45	17.86	21.65	46.70
	Maximum	89.58	87.88	74.72	80.77	82.30
Ensemble	Mean	83.01	53.68	59.45	55.72	73.33
	Std	3.36	11.30	14.77	11.25	7.93
	Minimum	77.08	31.79	33.65	34.11	60.28
	Maximum	91.67	80.68	85.83	76.92	87.45

The one-way ANOVA results in Table 3 show significant model effects for Accuracy ($F = 10.26$, $p = 2.64 \times 10^{-6}$), Recall ($F = 15.92$, $p = 2.65 \times 10^{-9}$), F1-score ($F = 7.77$, $p = 6.3 \times 10^{-5}$), and AUC ($F = 18.90$, $p = 8.31 \times 10^{-11}$). Precision is the exception ($F = 0.91$, $p = 0.43$), indicating that the observed differences in mean Precision are not statistically significant at $\alpha = 0.05$. Effect sizes calculated from the ANOVA sums of squares are approximately 0.136 for Accuracy, 0.014 for Precision, 0.196 for Recall, 0.106 for F1-score, and 0.224 for AUC. Thus, the strongest model effect is observed for AUC, followed by Recall, whereas the effect on Precision is negligible

The reported Tukey-Cramer analysis (Table 4) identifies no significant Euclidean-versus-ensemble F1-score difference, despite the ensemble's higher mean F1-score. Precision also lacks a significant omnibus ANOVA result ($p = 0.43$). Any reported omnibus difference in Accuracy, Recall, or AUC is across the four models and does not alone prove a material benefit over the strongest individual model. Furthermore, because models were applied to repeated partitions of the same dataset,

unpaired ANOVA assumptions merit scrutiny. Pairwise inferences should be rechecked using paired per-partition differences and an uncertainty interval; in the absence of those analyses the comparisons are descriptive rather than definitive evidence of production-level improvement.

Table 3
Analysis of Variance

Metrics	Sum of variance	Sum of squares	Mean squares	F	p-value	F-criteria
Accuracy	between	513.10	171.03	10.26	2.64×10^{-6}	2.65
	within	3267.49	16.67			
Precision	between	591.49	197.16	0.91	0.43	
	within	41559.51	212.04			
Recall	between	7801.73	2600.58	15.92	2.65×10^{-9}	
	within	32016.24	163.35			
F1 score	between	3211.33	1070.44	7.77	6.3×10^{-5}	
	within	27010.43	137.81			
AUC	between	2996.91	998.97	18.9	8.310×10^{-11}	
	within	10361.25	52.86			

The Tukey-Cramer results in Table 4 clarify which model pairs drive the significant omnibus effects. For Accuracy, all six pairwise mean differences exceed the stated critical difference. For Recall, all pairs are significant except Cosine versus Hamming. For F1-score, Euclidean versus Ensemble and Cosine versus Hamming are not significant, while the remaining comparisons are significant. For AUC, Cosine versus Hamming is the only nonsignificant pair. Because the Precision ANOVA is not significant, the Precision pairwise differences should not be interpreted as confirmatory evidence. Overall, the statistical analysis supports the ensemble advantage for Accuracy, Recall, and AUC, but it does not support a Precision advantage.

The present evidence has several important limitations. Only AISI 1045 specimens produced under one documented wet face-milling configuration were examined, and grade N3 has just two observations; robust estimation for this minority grade is therefore unavailable. No experiments changed tool wear, tool geometry, machine condition, coolant condition, batch, operator, or production environment, and no independent time-separated or machine-separated test dataset was used. Consequently, the reported metrics describe internal resampling only and cannot establish generalization to changing industrial conditions. Binary Hamming encoding also discards continuous information. Prediction latency, storage demands, preprocessing time, sensor-measurement time, and hardware feasibility were not assessed. In particular, the requirement for six tangential roughness measurements must be considered in any purported real-time workflow. Independent external validation, a cost-sensitive false-accept/false-reject analysis, and paired uncertainty analysis are necessary before claiming an operational advantage.

Table 4
Tukey-Cramer Ad Hoc Test

KNN Models Comparison	Absolute Mean Difference				
	Accuracy	Precision	Recall	F1 score	AUC
KNN-Euclidean vs KNN-Cosine	2.84	2.93	8.67	6.10	5.96
KNN-Euclidean vs KNN-Hamming	1.21	4.59	10.72	8.04	6.36
KNN-Euclidean vs Ensemble	1.52	1.34	4.57	1.51	2.68
KNN-Cosine vs KNN-Hamming	1.63	1.67	2.04	1.94	0.40
KNN-Cosine vs Ensemble	4.36	1.59	13.24	7.61	8.64
KNN-Hamming vs Ensemble	2.73	3.25	15.29	9.55	9.04

The study has four principal limitations. First, class N3 contains only two observations and cannot be evaluated reliably under repeated random hold-out. Second, the results constitute internal validation on one machining dataset and do not demonstrate transferability to other machines, tools, materials, or cutting-fluid conditions. Third, the binary representation required by Hamming distance may discard physically meaningful magnitude differences. Fourth, repeated random partitions are useful for estimating variability but do not replace an independent external test set. These limitations explain why the reported Accuracy is considerably higher than the class-sensitive Precision, Recall, and F1-score and should be considered when deploying the model for process control.

This study compared established Euclidean (k = 4), Cosine (k = 7), and binary-encoded Hamming (k = 14) KNN classifiers and their standard hard-voting ensemble for axial roughness-grade classification of wet face-milled AISI 1045 steel. Using three machining variables and six tangential roughness predictors, model performance was summarized over 50 repeated 80:20 hold-out partitions. The contribution is an application-specific analysis of distance-representation trade-offs rather than a novel classifier or ensemble algorithm.

4. Conclusion

Although the ensemble combines different neighborhood representations, its advantage over Euclidean KNN was modest: +1.52 percentage points in Accuracy, +1.51 in F1-score, +4.57 in Recall, and +2.68 in AUC, accompanied by −1.34 in Precision. A significantly higher F1-score relative to Euclidean KNN was not established by the reported Tukey-Cramer comparison. These results favor a conditional rather than universal recommendation: Euclidean KNN is the simpler baseline, whereas the ensemble may be considered where additional grade detection is more valuable than false alarms or inference overhead. The present study has not quantified these operational costs.

Reference [14] investigated the same AISI 1045 axial roughness classification problem using crisp/fuzzy discretized decision-tree models and decision-tree voting. The present study instead evaluates neighborhood geometry across three KNN variants and their hard vote. Earlier numerical benchmark differences (ensemble versus ensemble: +0.34 Accuracy, +4.26 Recall, +1.15 F1-score, +2.31 AUC, and −0.29 Precision points) are descriptive only: consistent splits, independent test specimens, and equivalence of dataset provenance have not been established. Similar experimental descriptions and overlapping authorship make disclosure of any shared dataset essential. The authors should explicitly document whether [14] and this study use fully shared, partially overlapping, or independent measurements before interpreting the comparison as evidence of methodological improvement.

Specify definitively whether the measurements in [14] were reused (fully or in part). Replace the uncertainty above with the verified dataset-provenance statement and document record overlap; if reused, frame this paper as a secondary analysis of previously published experimental data, not a new experiment. Also supply paired resampling outputs and timing measurements if stronger inferential or industrial claims are desire.

The primary limitations are the small and imbalanced dataset (including two N3 observations), dependence on a single material and machining setting, possible information loss in binary Hamming encoding, and lack of independent production samples or implementation measurements. No real-time capability, cross-machine generalizability, or industrial quality-cost advantage has been demonstrated. Future work should disclose dataset provenance relative to [14], acquire independent test batches under tool-wear and machining-condition changes, apply leakage-free nested or stratified validation, calculate paired confidence intervals for ensemble-versus-Euclidean differences,

and measure prediction latency, memory, and inspection-cycle costs. Those results are needed before recommending routine industrial deployment.

The improvement achieved by the ensemble can be explained by the complementary error characteristics of the three base classifiers. Euclidean KNN preserves normalized magnitude information and supplies the most stable predictions, Cosine KNN evaluates directional similarity, and Hamming KNN introduces a discretized mismatch representation. Through majority voting, errors from one classifier can be corrected when the other two classifiers agree. However, the relatively modest gain over Euclidean KNN indicates that the ensemble remains constrained by the weaker component models and the limited representation of minority roughness grades.

Benchmarking against the decision-tree models reported in [14] further demonstrates the competitiveness of the proposed approach. Compared with the ensemble of multiple decision trees, the KNN ensemble improved Accuracy by 0.34 percentage points, Recall by 4.26 points, F1-score by 1.15 points, and AUC by 2.31 points, while Precision decreased by 0.29 points. Relative to the ID3 decision tree, the proposed ensemble improved Accuracy by 4.51 points, Recall by 12.09 points, F1-score by 0.33 points, and AUC by 7.05 points, although Precision was 13.57 points lower. Thus, the proposed method is particularly advantageous in terms of Recall and AUC, but it does not outperform the benchmark methods across every metric.

The main limitations of this study are the severe underrepresentation of class N3, the use of a single machining dataset, and the information loss caused by binary conversion for the Hamming distance. Future work should collect additional minority-class observations, apply stratified or nested validation, investigate weighted and distance-weighted voting schemes, and verify the model using independent machining experiments.

Acknowledgement
This work was supported by Universitas Sriwijaya, Fiscal Year 2025. In accordance with the Rector's Decree No.0029/UN9/SK.LP2M.PT/2025, September 17, 2025 (Profession Research scheme).

Conflict of Interest Statement
The authors declare that there is no conflict of interest regarding the publication of this paper. No financial support, grants, or other forms of compensation were received that could have influenced the outcomes of this work. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions Statement
M. A. Ade Saputra: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft.
Irsyadi Yani: Conceptualization, Methodology, Validation, Supervision, Writing – review & editing.
Yulia Resti: Methodology, Formal analysis, Validation, Statistical analysis, Writing – review & editing.
Dewi Puspitasari: Investigation, Validation, Supervision, Writing – review & editing.
Ismail Thamrin: Investigation, Resources, Validation, Writing – review & editing.
Muhammad Yanis: Conceptualization, Investigation, Resources, Supervision, Writing – review & editing.
All authors have reviewed and approved the final version of the manuscript and agreed to be accountable for their respective contributions to the work.

Data Availability Statement

All data generated or analyzed during this study are included in this published article. Additional datasets are available from the corresponding author upon reasonable request. Where applicable, publicly available datasets used in the study are cited in the references.

Ethics Statement

This study was conducted in accordance with the ethical standards of the institutional and/or national research committee. Ethical approval was obtained where required, and informed consent was obtained from all participants involved in the research.

References

[1] I. Yani, I. Thamrin, D. Puspitasari, Y. Resti, and D. Adanta, "Plastic-Type Classification for Sorting System Based on Digital Image Using Multinomial Logistic Regression with k-Fold Cross Validation," *Eng. Lett.*, vol. 33, no. 4, pp. 1081–1088, 2025.

[2] E. O. Kiyak and B. Ghasemkhani, "High-Level K-Nearest Neighbors (HLKNN): A Supervised," *Electronics*, vol. 12, pp. 1–20, 2023.

[3] J. Chen, J. Lin, M. Zhang, and Q. Lin, "Predicting Surface Roughness in Turning Complex-Structured Workpieces Using Vibration-Signal-Based Gaussian Process Regression," *Sensors*, vol. 24, no. 7, 2024, doi: 10.3390/s24072117.

[4] V. Dubey, A. K. Sharma, and D. Y. Pimenov, "Prediction of Surface Roughness Using Machine Learning Approach in MQL Turning of AISI 304 Steel by Varying Nanoparticle Size in the Cutting Fluid," *Lubricants*, vol. 10, no. 5, 2022, doi: 10.3390/lubricants10050081.

[5] P. Wu, M. Xia, and L. Hu, "A Real Time Condition Based Sustainable Maintenance Method for Milling Process," *J. Clean. Prod.*, vol. 434, no. November 2023, p. 140284, 2024, doi: 10.1016/j.jclepro.2023.140284.

[6] N. Ali, D. Neagu, and P. Trundle, "Evaluation of K-Nearest Neighbour Classifier Performance for Heterogeneous Data Sets," *SN Appl. Sci.*, vol. 1, no. 12, pp. 1–15, 2019, doi: 10.1007/s42452-019-1356-9.

[7] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-Nearest Neighbor Algorithm: A Comprehensive Review and Performance Analysis of Modifications," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00973-y.

[8] M. Bansal, A. Goyal, and A. Choudhary, "A Comparative Analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory Algorithms in Machine Learning," *Decis. Anal. J.*, vol. 3, no. November 2021, p. 100071, 2022, doi: 10.1016/j.dajour.2022.100071.

[9] S. H. Noh, "Comparing the Performance of Corporate Bankruptcy Prediction Models Based on Imbalanced Financial Data," *Sustain.* , vol. 15, no. 6, 2023, doi: 10.3390/su15064794.

[10] L. Y. Hu, M. W. Huang, S. W. Ke, and C. F. Tsai, "The Distance Function Effect on K-Nearest Neighbor Classification for Medical Datasets," *Springerplus*, vol. 5, no. 1, 2016, doi: 10.1186/s40064-016-2941-7.

[11] A. Berti, A. Bernasconi, G. M. Del Corso, and R. Guidotti, "The Role of Encodings and Distance Metrics for the Quantum Nearest Neighbor," *Quantum Mach. Intell.*, vol. 6, no. 2, 2024, doi: 10.1007/s42484-024-00197-6.

[12] I. Arora, N. Khanduja, and M. Bansal, "Effect of Distance Metric and Feature Scaling on KNN Algorithm While Classifying X-Rays," *CEUR Workshop Proc.*, vol. 3176, pp. 61–75, 2022.

[13] C. Feng, B. Zhao, X. Zhou, X. Ding, and Z. Shan, "An Enhanced Quantum K-Nearest Neighbor Classification Algorithm Based on Polar Distance," *Entropy*, vol. 25, no. 1, pp. 1–12, 2023, doi: 10.3390/e25010127.

[14] Y. Resti, I. Yani, D. Puspitasari, I. Thamrin, and M. A. Ade Saputra, "Ensemble Method of Multiple Decision Trees with Crisp and Fuzzy Discretization for Axial Surface Roughness Prediction," *Sci. Technol. Indones.*, vol. 10, no. 3, pp. 678–689, 2025, doi: 10.26554/sti.2025.10.3.678-689.

[15] A. Balasuadhakar, S. Thirumalai Kumaran, and M. Uthayakumar, "Machine Learning Prediction of Surface Roughness in Sustainable Machining of AISI H11 Tool Steel," *Smart Mater. Manuf.*, vol. 3, no. February, p. 100075, 2025, doi: 10.1016/j.smmf.2025.100075.

[16] M. Z. Zahaf and M. Benghersallah, "Surface Roughness and Vibration Analysis in End Milling of Annealed and Hardened Bearing Steel," *Meas. Sensors*, vol. 13, no. September 2020, p. 100035, 2021, doi: 10.1016/j.measen.2020.100035.

[17] H. Qu, J. Xu, Z. Li, D. Wei, and F. Wang, "Effects of Embedded Distance Measurements Interacting with Modeling Approaches on Empirical Dynamical Model Predictions," *Ecol. Indic.*, vol. 146, no. December 2022, p. 109895, 2023, doi: 10.1016/j.ecolind.2023.109895.

[18] M. Zhao, J. Chen, and Y. Li, "A Review of Anomaly Detection Techniques Based on Nearest Neighbor," *Proc. 2018 Int. Conf. Comput. Model. Simul. Algorithm (CMSA 2018)*, vol. 151, no. Cmsa, pp. 290–292, 2018, doi: 10.2991/cmsa-18.2018.65.

[19] S. Nayak, M. Bhat, N. V. S. Reddy, and B. A. Rao, "Study of Distance Metrics on k - Nearest Neighbor Algorithm for Star Categorization," *J. Phys. Conf. Ser.*, vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012004.

[20] Y. Ruan, X. Xue, H. Liu, J. Tan, and X. Li, "Quantum Algorithm for K-Nearest Neighbors Classification Based on the Metric of Hamming Distance," *Int. J. Theor. Phys.*, vol. 56, no. 11, pp. 3496–3507, 2017, doi: 10.1007/s10773-017-3514-4.

[21] I. K. Nti, O. Nyarko-Boateng, and J. Aning, "Performance of Machine Learning Algorithms with Different K Values in K-Fold CrossValidation," *Int. J. Inf. Technol. Comput. Sci.*, vol. 13, no. 6, pp. 61–71, 2021, doi: 10.5815/ijitcs.2021.06.05.

[22] W. Chandra, B. Suprihatin, and Y. Resti, "Median-KNN Regressor-SMOTE-Tomek Links for Handling Missing and Imbalanced Data in Air Quality Prediction," *Symmetry (Basel).*, vol. 15, no. 4, p. 887, Apr. 2023, doi: 10.3390/sym15040887.

[23] A. İ. Çetin and A. H. Büyüklü, "A New Approach to K-Nearest Neighbors Distance Metrics on Sovereign Country Credit Rating," *Kuwait J. Sci.*, vol. 52, no. 1, 2025, doi: 10.1016/j.kjs.2024.100324.

[24] A. A. Amer, S. D. Ravana, and R. A. A. Habeeb, "Effective K-Nearest Neighbor Models for Data Classification Enhancement," *J. Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01137-2.

[25] A. Tzotzis, P. Maropoulos, and P. Kyratsis, "A Dynamic Surface Roughness Prediction System Based on Machine Learning for the 3D-Printed Carbon-Fiber-Reinforced-Polymer (CFRP) Turning," *J. Intell. Manuf.*, vol. 7075, 2025, doi: 10.1007/s10845-025-02602-8.

[26] I. Thamrin, "Analysis of the Effect of the Cutting Speed on Specific Cutting Energy in Turning Process," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 620, no. 1, 2019, doi: 10.1088/1757-899X/620/1/012117.

[27] S. H. Rukmawan, F. R. Aszhari, Z. Rustam, and J. Pandelaki, "Cerebral Infarction Classification Using the K-Nearest Neighbor and Naive Bayes Classifier," *J. Phys. Conf. Ser.*, vol. 1752, no. 1, 2021, doi: 10.1088/1742-6596/1752/1/012045.

[28] B. M. Alsafy, Z. M. Aydam, and W. K. Mutlag, "Multiclass Classification : A Review," *Int. J. Adv. Engeineering Technol. Innov. Sci.*, vol. 5, no. 3, pp. 01–10, 2014.

[29] R. Mussabayev, "Optimizing Euclidean Distance Computation," *Mathematics*, vol. 12, no. 23, pp. 1–36, 2024, doi: 10.3390/math12233787.