Quest Journals
Journal of Medical and Dental Science Research
Volume 13~ Issue 2 (February 2026) pp: 32-43
ISSN(Online): 2394-076X ISSN (Print):2394-0751
www.questjournals.org

Research Paper

Comparison of AI-Assisted and Conventional Methods for Predicting the Difficulty of Mandibular Third Molar Surgery

Sanjoy Dutta[1], Sayan Chattopadhyay[2], Rachita Arora[3]
[1]*Professor, Department of Oral & Maxillofacial Surgery, North Bengal Dental College & Hospital, Darjeeling, West Bengal, India*
[2]*Associate Professor, Department of Oral Medicine & Radiology, Dr. R. Ahmed Dental College & Hospital, Kolkata, West Bengal, India*
[3]*Associate Professor, Department of Pediatric & Preventive Dentistry, Dr. R. Ahmed Dental College & Hospital, Kolkata, West Bengal, India*
Corresponding author: *Sanjoy Dutta*

Abstract
Background:
Predicting the surgical difficulty of impacted mandibular third molar removal is important for treatment planning and anticipating operative challenges. Conventional radiographic classifications provide useful information but may have limitations in accurately estimating surgical complexity. Artificial intelligence (AI) may offer an alternative approach by identifying radiographic patterns associated with difficult extractions.
Aim:
To compare the accuracy of an AI-assisted approach with conventional radiographic assessment for predicting the surgical difficulty of impacted mandibular third molar removal
Materials and Methods:
A comparative observational study was conducted in 100 patients aged 18 years or older who underwent surgical removal of an impacted mandibular third molar between September and December 2025. Preoperative panoramic radiographs were assessed using Winter's classification and the Pell–Gregory classification for tooth angulation, ramus relationship, and depth of impaction. Conventional prediction was based on a predefined modified Winter–Pell–Gregory scoring system. AI-assisted prediction was performed using a ResNet50-based deep learning model and categorized cases as having low, moderate, or high surgical difficulty. Observed surgical difficulty and operative time were used as reference outcomes. Predictive performance was assessed using accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, Cohen's kappa, and receiver operator curve (ROC) analysis and the area under the ROC curve (AUC).
Results:
The mean age of the study population was 25.8 ± 5.1 years, with a slight male predominance (56%). High surgical difficulty was observed in 42% of cases, and mean operative time increased progressively from low- to high-difficulty procedures. The AI-assisted method demonstrated significantly greater accuracy than the conventional approach (89.0% vs. 76.0%; p = 0.012) and higher sensitivity (90.5% vs. 73.8%; p = 0.041). The AI-assisted method also showed a higher F1 score (0.884 vs. 0.747) and superior agreement with observed surgical difficulty (κ = 0.77 vs. 0.51). ROC analysis demonstrated significantly better discrimination with AI-assisted prediction (AUC 0.93; 95% CI: 0.88–0.97) compared with conventional assessment (AUC 0.81; 95% CI: 0.73–0.88; p = 0.006). Age, Pell–Gregory Class III ramus relationship, and Position C depth were independently associated with greater surgical difficulty.
Conclusion:
The ResNet50-based AI-assisted approach demonstrated superior predictive performance compared with conventional radiographic assessment in estimating the surgical difficulty of impacted mandibular third molar removal. Its higher accuracy, sensitivity, agreement, and discriminatory ability suggest that AI-assisted radiographic analysis may serve as a useful adjunct for preoperative surgical planning and risk assessment. Further validation using larger, multicentre datasets is warranted before routine clinical implementation.

Keywords: ***Artificial Intelligence, Deep Learning, Impacted Mandibular Third Molar, Radiographic Assessment, Surgical Difficulty***

*Received 17 Feb., 2026; **R**evised 24 Feb., 2026; **A**ccepted 27 Feb., 2026 © The author(s) 2026.*
P**ublished with open access at **www.questjournals.org

I. Introduction

Surgical removal of impacted mandibular third molars is a routine procedure in oral and maxillofacial practice, yet the degree of operative difficulty can vary considerably between patients. Difficult extraction may be associated with prolonged operating time, increased tissue trauma, greater postoperative morbidity, and a higher likelihood of surgical complications. Consequently, reliable prediction of surgical difficulty before treatment is important for appropriate case selection, informed patient counselling, operative planning, and allocation of surgical expertise.

Traditional prediction of mandibular third molar surgical difficulty has primarily relied on clinical examination and radiographic classification systems. The depth of impaction, relationship of the tooth to the ascending ramus, and its angulation are commonly assessed using established classifications such as the Pell–Gregory and Winter systems. Although these methods provide a practical framework for describing tooth position, they may not adequately reflect the full range of anatomical and clinical variables that contribute to extraction complexity [1].

Research has demonstrated that surgical difficulty is influenced by several interacting patient- and tooth-related characteristics. Factors such as patient age, body habitus, depth of impaction, degree of bony coverage, root morphology, and the accessibility of the tooth may affect operative time and technical complexity [2]. In addition, anatomical characteristics and operative conditions may have a greater influence on extraction difficulty than demographic variables alone [3]. These observations suggest that third molar surgery should be regarded as a multifactorial procedure in which reliance on a limited number of conventional radiographic parameters may result in incomplete risk estimation.

Advances in digital imaging have provided additional opportunities for detailed preoperative evaluation. Cone-beam computed tomography (CBCT) enables three-dimensional (3D) visualization of the impacted tooth and surrounding structures, including root configuration, cortical bone, mandibular canal position, and the available osseous space. Incorporating such information may improve anatomical characterization and provide additional measurable variables for predictive assessment.

Despite these advances, estimation of surgical difficulty can still be influenced by the experience and subjective interpretation of the clinician. When numerous clinical and radiographic variables must be considered simultaneously, differences in individual assessment may occur. Artificial intelligence (AI) and machine-learning techniques offer a potential solution by enabling the simultaneous analysis of multiple variables and recognition of complex patterns that may not be readily apparent through conventional assessment.

Recent AI-based approaches have demonstrated promising potential for predicting mandibular third molar difficulty from radiographic images [4]. Deep-learning systems can analyze panoramic radiographs and classify cases according to anticipated surgical complexity, providing a standardized approach to image interpretation. Multimodal machine-learning models that combine clinical information with CBCT-derived anatomical characteristics have further expanded this concept and may offer improved predictive performance by incorporating complementary sources of information [5].

Automated image analysis may also facilitate objective quantification of radiographic characteristics such as tooth inclination, impaction depth, and the relationship between the third molar and mandibular canal [6]. By converting image-based observations into reproducible measurements, AI-assisted systems may reduce variability associated with purely subjective assessment and facilitate more consistent preoperative decision-making.

The relevance of AI is particularly evident because surgical difficulty rarely depends on a single anatomical characteristic. Instead, it may arise from specific combinations of tooth position, root morphology, surrounding bone, patient characteristics, and anticipated operative requirements. Machine-learning models are capable of integrating these diverse variables and identifying relationships among them, potentially allowing a more individualized prediction of surgical difficulty than conventional scoring approaches.

Nevertheless, AI should be considered a decision-support technology rather than a substitute for clinical expertise. Its reliability depends on the quality and representativeness of the training dataset, consistency of imaging protocols, accurate definition of surgical outcomes, and appropriate external validation. Differences in patient populations, radiographic equipment, operator expertise, and surgical techniques may also affect the generalizability of AI models. Therefore, direct comparison between AI-assisted prediction and established conventional assessment is necessary to determine whether AI provides meaningful improvement in the preoperative prediction of mandibular third molar surgical difficulty.

Therefore, the present study was designed to evaluate the predictive performance of an AI-based approach against established clinical and radiographic assessments. The study aims to determine whether AI can provide a more consistent and informative estimation of surgical difficulty, thereby contributing to improved preoperative planning and individualized management of impacted mandibular third molars.

II. Materials & Methods

Study Design and Population

A comparative observational study was conducted at a single tertiary care centre from September to December 2025 to evaluate the performance of an AI-assisted approach in predicting the surgical difficulty of impacted mandibular third molar removal and to compare its performance with conventional radiographic assessment. The study included 100 patients aged 18 years or older who underwent surgical removal of an impacted mandibular third molar.

Inclusion Criteria

- Patients aged 18 years or older
- Patients requiring surgical removal of an impacted mandibular third molar
- Presence of a radiographically confirmed impacted mandibular third molar
- Pre operative radiographs of sufficient quality to evaluate Winter’s angulation and Pell–Gregory ramus relationship and depth of impaction
- Patients who provided informed consent for participation in the study
- Complete clinical and operative records, including operative time and assessment of surgical difficulty

Exclusion Criteria

- Patients younger than 18 years
- Patients without a radiographically confirmed impacted mandibular third molar
- Radiographs with poor image quality, artifacts, or inadequate visualization of the third molar region
- Patients with pathological lesions or significant anatomical abnormalities associated with the mandibular third molar region
- Previously surgically treated or partially removed mandibular third molar
- Patients with incomplete clinical, radiographic, or operative records
- Patients in whom the operative time or surgical difficulty could not be reliably documented
- Cases in which radiographic findings could not be classified according to the predefined Winter’s or Pell–Gregory criteria

Radiographic Assessment

Preoperative panoramic radiographs were evaluated for features associated with the difficulty of mandibular third molar removal. The angulation of the impacted tooth was categorized according to Winter’s classification as mesioangular, vertical, horizontal, or distoangular. The relationship of the impacted tooth to the mandibular ramus and its depth relative to the occlusal plane were assessed according to the Pell–Gregory classification. For ramus relationship, cases were classified as Class I, Class II, or Class III, while the depth of impaction was categorized as Position A, B, or C. These radiographic parameters were used for conventional assessment, AI-assisted prediction, and subgroup analysis.

Surgical Difficulty Scoring

Surgical difficulty was assessed using a modified Winter–Pell–Gregory scoring system based on the three principal radiographic variables: tooth angulation, ramus relationship, and depth of impaction**.** Each radiographic parameter was assigned a numerical score according to its anticipated contribution to surgical complexity. Mesioangular and vertical impactions were assigned lower scores, whereas horizontal and distoangular orientations received higher scores**.** Similarly, Pell–Gregory Class I and Position A represented lower difficulty, while Class III and Position C represented greater difficulty. The individual scores were summed to obtain a total radiographic difficulty score. Based on the predefined score ranges, cases were categorized into low, moderate, or high surgical difficulty. The scoring system was established before analysis and was applied consistently to all cases.

Conventional Prediction

The conventional method used the predefined radiographic scoring criteria to estimate the anticipated difficulty of third molar removal. The assessment incorporated Winter’s angulation, Pell–Gregory ramus relationship, and Pell–Gregory depth of impaction. The predicted difficulty was classified as low, moderate, or high.

AI-Assisted Prediction

The AI-assisted assessment was performed using a ResNet50-based deep learning model. Preoperative radiographic images were standardized before analysis to reduce variations related to image orientation, quality, and magnification. The ResNet50 model analyzed the radiographic characteristics of the impacted mandibular third molar and generated a prediction of surgical difficulty. The output was categorized into low, moderate, or high difficulty, corresponding to the predefined clinical classification. The radiographic dataset was used to develop and evaluate the model according to a predefined training and testing protocol. Model performance was subsequently compared with the conventional radiographic method using the observed surgical difficulty as the reference outcome.

Assessment of Observed Surgical Difficulty

The actual difficulty of third molar removal was determined from the predefined difficulty classification and intraoperative findings. Operative time was recorded in minutes from initiation of the surgical procedure until completion of removal of the impacted third molar. Based on the predefined criteria, cases were categorized as low, moderate, or high difficulty. The observed difficulty category served as the reference standard for evaluating both prediction methods.

Outcome Measures

The primary outcome was the comparative accuracy of the AI-assisted and conventional methods in predicting surgical difficulty. Secondary outcomes included agreement between predicted and observed difficulty, prediction accuracy across different radiographic subgroups, operative time according to difficulty category, and clinical and radiographic factors associated with high surgical difficulty. Prediction performance was evaluated using accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 score. Overall discriminatory performance was assessed using receiver operator curve (ROC) analysis and the area under the ROC curve (AUC).

Statistical Analysis

All statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS), IBM SPSS Statistics for Windows, version 26.0 (IBM Corp., Armonk, NY, USA). Appropriate descriptive and inferential statistical methods were applied according to the type and distribution of the data. Continuous variables such as age were expressed as mean ± standard deviation (SD), while categorical variables were summarized as frequencies and percentages. The distribution of continuous variables was assessed before analysis. Prediction accuracy between the AI-assisted and conventional methods was compared using appropriate paired categorical-data analysis. Agreement between predicted and observed surgical difficulty was assessed using Cohen's kappa (κ) with 95% confidence interval (CI). Differences in categorical variables were assessed using the Chi-square test or Fisher's exact test, as appropriate. All statistical tests were two-sided, and a p-value < 0.05 was considered statistically significant.

III. Results

Table 1 & Figure 1 summarize the demographic characteristics of the study population. A total of 100 patients undergoing surgical removal of impacted mandibular third molars were included. The mean age was 25.8 ± 5.1 years. Participants aged 18–24 years constituted the largest age group (43.0%), followed by those aged 25–29 years (31.0%) and ≥30 years (26.0%). Males accounted for 56.0% of the participants, compared with 44.0% females, demonstrating a slight male predominance.

Table 1: Study population and demographic characteristics

Variable	n (%) / Mean ± SD
Age (years)	25.8 ± 5.1
18–24 years	43 (43.0%)
25–29 years	31 (31.0%)
≥30 years	26 (26.0%)
Male	56 (56.0%)
Female	44 (44.0%)

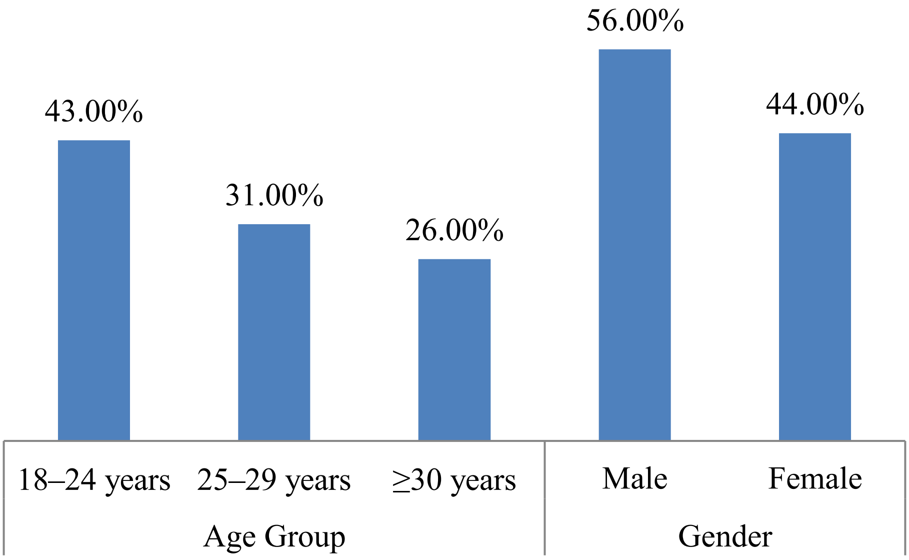

Figure 1: Age and gender distribution of the study participants

Table 2 & Figure 2 summarize the radiographic characteristics of the impacted mandibular third molars. Based on Winter's classification, mesioangular impaction was the most prevalent pattern, observed in 46 (46.0%) cases, followed by vertical impaction in 29 (29.0%) cases. Horizontal and distoangular orientations were identified in 16 (16.0%) and 9 (9.0%) cases, respectively. According to the Pell–Gregory classification of ramus relationship, Class II was predominant, accounting for 52 (52.0%) cases, followed by Class I in 34 (34.0%) and Class III in 14 (14.0%) cases. With respect to the depth of impaction, Position B was most frequently observed (47; 47.0%), followed by Position A (38; 38.0%) and Position C (15; 15.0%) cases.

Table 2: Radiographic characteristics of mandibular third molars

Radiographic parameter	Category	n (%)
Winter's angulation	Mesioangular	46 (46.0%)
	Vertical	29 (29.0%)
	Horizontal	16 (16.0%)
	Distoangular	9 (9.0%)
Pell–Gregory ramus	Class I	34 (34.0%)
	Class II	52 (52.0%)
	Class III	14 (14.0%)
Pell–Gregory depth	Position A	38 (38.0%)
	Position B	47 (47.0%)
	Position C	15 (15.0%)

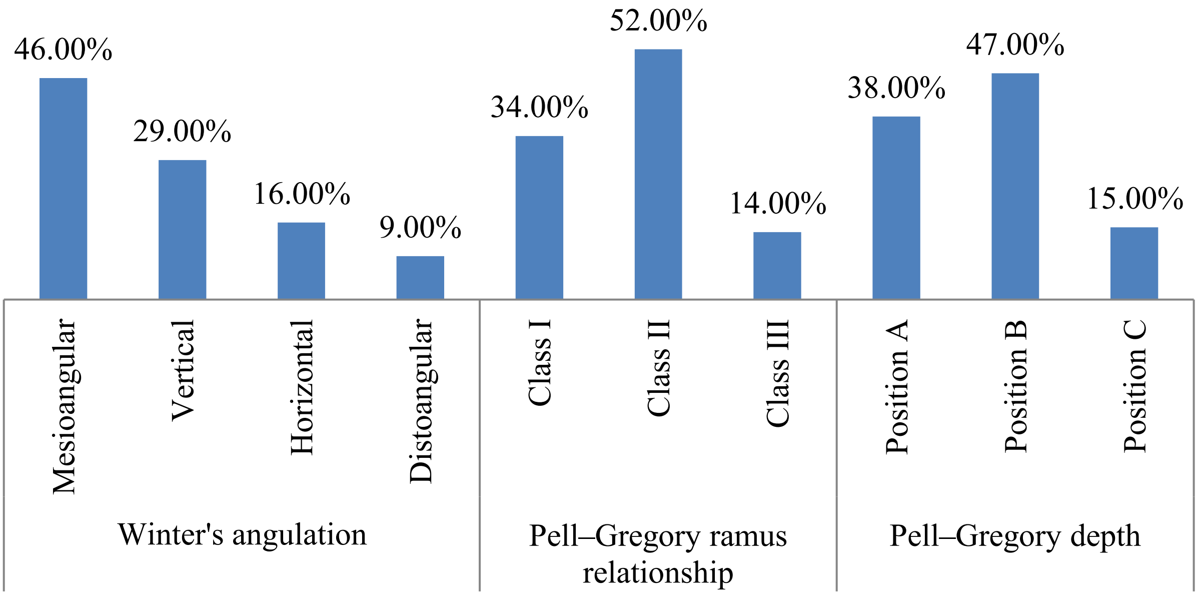

Figure 2: Distribution of impacted mandibular third molars according to Winter's and Pell–Gregory classifications

Table 3 & Figure 3 present the distribution of cases according to surgical difficulty and the corresponding operative time. Based on the predefined difficulty assessment, 31 (31.0%) cases were classified as low difficulty, 27 (27.0%) as moderate difficulty, and 42 (42.0%) as high difficulty. Operative time showed a progressive increase with greater surgical difficulty, with mean durations of 18.4 ± 4.2 minutes, 27.1 ± 5.1 minutes, and 41.8 ± 8.7 minutes for low-, moderate-, and high-difficulty cases, respectively. The overall mean operative time was 30.9 ± 11.2 minutes.

Table 3: Observed surgical difficulty

Difficulty category	n (%)	Mean operative time (min)
Low difficulty	31 (31.0%)	18.4 ± 4.2
Moderate difficulty	27 (27.0%)	27.1 ± 5.1
High difficulty	42 (42.0%)	41.8 ± 8.7
Overall	100 (100.0%)	30.9 ± 11.2

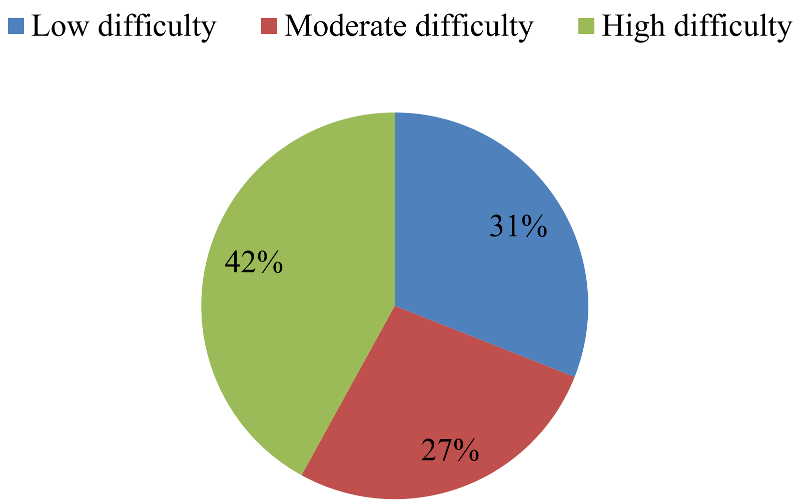

Figure 3: Distribution of surgical difficulty and mean operative time according to difficulty level

Table 4 & Figure 4 compare the predictive performance of the AI-assisted and conventional methods for identifying surgical difficulty. The AI-assisted method achieved significantly higher overall accuracy than the conventional method (89.0% vs. 76.0%; p = 0.012). Sensitivity was also significantly greater with AI-assisted prediction (90.5% vs. 73.8%; p = 0.041). Although specificity was higher for the AI-assisted method than for the conventional approach (87.9% vs. 77.6%), the difference was not statistically significant (p = 0.098). The AI-assisted method demonstrated a PPV of 86.4% and a NPV of 91.2%, compared with 75.6% and 75.9%, respectively, for the conventional method. The F1 score was also higher with AI-assisted prediction (0.884 vs. 0.747). Furthermore, the AI-assisted method showed superior discriminative ability, with a higher ROC-AUC than the conventional method (0.93 vs. 0.81; p = 0.006).

Table 4: Comparison of AI-assisted and conventional prediction

Performance measure	AI-assisted method	Conventional method	p-value
Accuracy	89.0%	76.0%	0.012*
Sensitivity	90.5%	73.8%	0.041*
Specificity	87.9%	77.6%	0.098
PPV	86.4%	75.6%	—
NPV	91.2%	75.9%	—
F1 score	0.884	0.747	—
ROC-AUC	0.93	0.81	0.006*

*p < 0.05 considered statistically significant

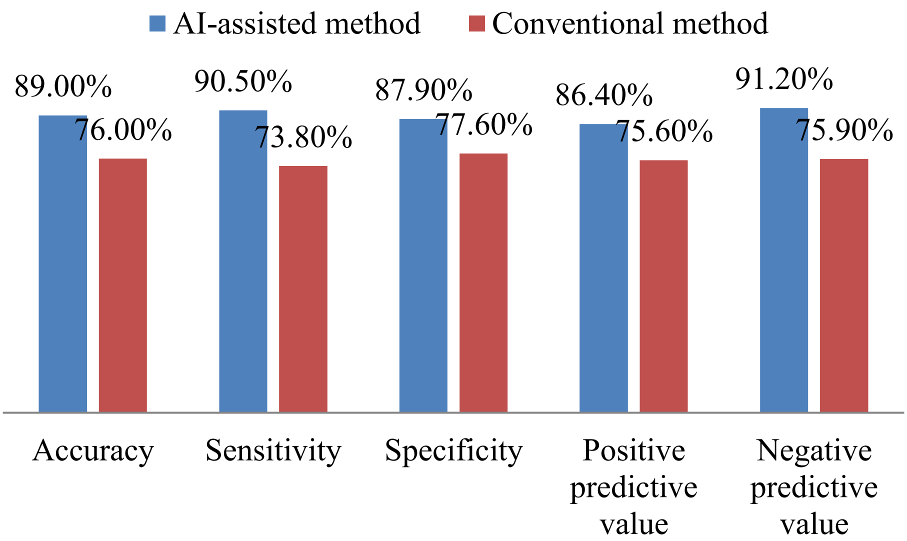

Figure 5: Agreement between predicted and observed surgical difficulty for AI-assisted and conventional methods

Table 5 shows the agreement between the predicted and observed levels of surgical difficulty. Cohen's kappa analysis demonstrated substantial agreement for the AI-assisted method, with a κ value of 0.77 (95% CI: 0.64–0.90; $p < 0.001$). In comparison, the conventional method showed moderate agreement, with a κ value of 0.51 (95% CI: 0.36–0.66; $p < 0.001$). Overall, the AI-assisted method demonstrated a higher level of concordance with the observed surgical difficulty than the conventional approach.

Table 5: Agreement between predicted and observed surgical difficulty

Method	Cohen's κ	95% CI	p-value	Interpretation
AI-assisted	0.77	0.64–0.90	<0.001*	Substantial agreement
Conventional	0.51	0.36–0.66	<0.001*	Moderate agreement

Table 6 & Figure 5 summarize the classification results obtained using the AI-assisted and conventional approaches. The AI-assisted method correctly classified surgical difficulty in 89 cases (89.0%), whereas the conventional method achieved correct classification in 76 cases (76.0%). False-positive results were recorded in 6 cases (6.0%) with AI-assisted prediction and 11 cases (11.0%) with the conventional approach. Similarly, false-negative results occurred in 5 cases (5.0%) and 13 cases (13.0%), respectively. Overall, the AI-assisted approach produced fewer classification errors and showed superior accuracy in predicting surgical difficulty.

Table 6: Distribution of correct and incorrect predictions

Prediction outcome	AI-assisted n (%)	Conventional n (%)
Correct prediction	89 (89.0%)	76 (76.0%)
False positive	6 (6.0%)	11 (11.0%)
False negative	5 (5.0%)	13 (13.0%)
Total	100 (100.0%)	100 (100.0%)

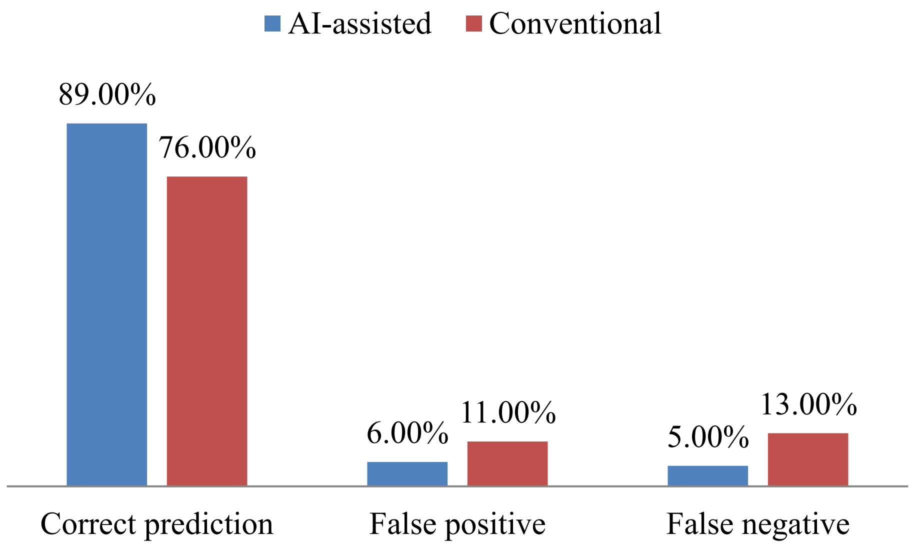

Figure 5: Comparison of classification outcomes between AI-assisted and conventional methods

Table 7 & Figure 6 present the subgroup analysis of prediction accuracy across different radiographic characteristics. For ramus relationship, the AI-assisted method achieved accuracies of 91.2%, 89.4%, and 85.7% for Classes I, II, and III, respectively, compared with 82.4%, 76.9%, and 57.1% for the conventional method. The difference was statistically significant for Class II ($p = 0.041$) and Class III ($p = 0.031$), but not for Class I ($p = 0.214$). For impaction depth, AI-assisted prediction achieved accuracies of 92.1%, 89.4%, and 80.0% for Positions A, B, and C, respectively, whereas the corresponding values for the conventional method were 84.2%, 74.5%, and 53.3%. Significant differences were observed for Position B ($p = 0.049$) and Position C ($p = 0.048$), while the difference for Position A was not statistically significant ($p = 0.286$). Across Winter's angulation categories, the AI-assisted method consistently demonstrated higher accuracy than the conventional approach for mesioangular, vertical, horizontal, and distoangular impactions. The difference reached statistical significance for horizontal impaction (81.3% vs. 56.3%; $p = 0.047$), whereas no significant differences were observed for mesioangular, vertical, or distoangular impactions.

Table 7: Comparison according to radiographic difficulty factors

Subgroup	AI accuracy	Conventional accuracy	p-value
Class I	91.2%	82.4%	0.214
Class II	89.4%	76.9%	0.041*
Class III	85.7%	57.1%	0.031*
Position A	92.1%	84.2%	0.286
Position B	89.4%	74.5%	0.049*
Position C	80.0%	53.3%	0.048*
Mesioangular	91.3%	80.4%	0.112
Vertical	93.1%	86.2%	0.418
Horizontal	81.3%	56.3%	0.047*
Distoangular	88.9%	66.7%	0.295

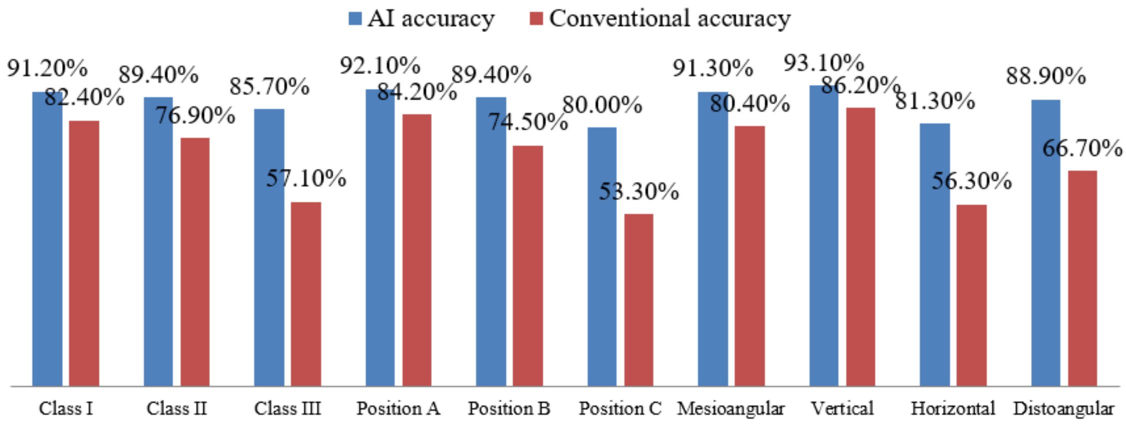

Figure 6: Subgroup comparison of AI-assisted and conventional prediction accuracy across radiographic characteristics

Table 8 presents the results of the multivariable logistic regression analysis performed to identify factors independently associated with high surgical difficulty. Age was significantly associated with greater odds of high surgical difficulty, with the odds increasing by 6% for each additional year of age (adjusted OR = 1.06; 95% CI: 1.01–1.12; p = 0.021). Using Class I ramus relationship as the reference category, Class II showed higher odds of high surgical difficulty, although the association was not statistically significant (adjusted OR = 1.84; 95% CI: 0.89–3.80; p = 0.097). Class III was independently associated with high surgical difficulty (adjusted OR = 3.72; 95% CI: 1.41–9.81; p = 0.008).

Compared with Position A, Position B was not significantly associated with high surgical difficulty (adjusted OR = 1.76; 95% CI: 0.86–3.61; p = 0.119), whereas Position C showed a significant association (adjusted OR = 4.15; 95% CI: 1.47–11.72; p = 0.007). Horizontal impaction was associated with higher odds of high surgical difficulty compared with mesioangular impaction, but the association did not reach statistical significance (adjusted OR = 2.31; 95% CI: 0.96–5.55; p = 0.062).

Table 8: Association of clinical/radiographic factors with high surgical difficulty

Predictor	Adjusted OR	95% CI	p-value
Age (per year)	1.06	1.01–1.12	0.021*
Class II vs I	1.84	0.89–3.80	0.097
Class III vs I	3.72	1.41–9.81	0.008*
Position B vs A	1.76	0.86–3.61	0.119
Position C vs A	4.15	1.47–11.72	0.007*
Horizontal vs mesioangular	2.31	0.96–5.55	0.062

Table 9 & Figure 7 summarize the ROC analysis comparing the discriminatory performance of the AI-assisted and conventional methods. The AI-assisted method demonstrated excellent discriminatory ability, with an AUC of 0.93 (95% CI: 0.88–0.97), whereas the conventional method showed comparatively lower discriminatory performance, with an AUC of 0.81 (95% CI: 0.73–0.88). At the optimal cut-off value, the AI-assisted method achieved a sensitivity of 90.5% and specificity of 87.9%, compared with 73.8% and 77.6%, respectively, for the conventional method. The difference in AUCs between the two methods was statistically significant (p = 0.006), indicating significantly better overall discrimination with the AI-assisted approach.

Table 9: ROC analysis

Method	AUC	95% CI	Optimal sensitivity	Optimal specificity
AI-assisted	0.93	0.88–0.97	90.5%	87.9%
Conventional	0.81	0.73–0.88	73.8%	77.6%

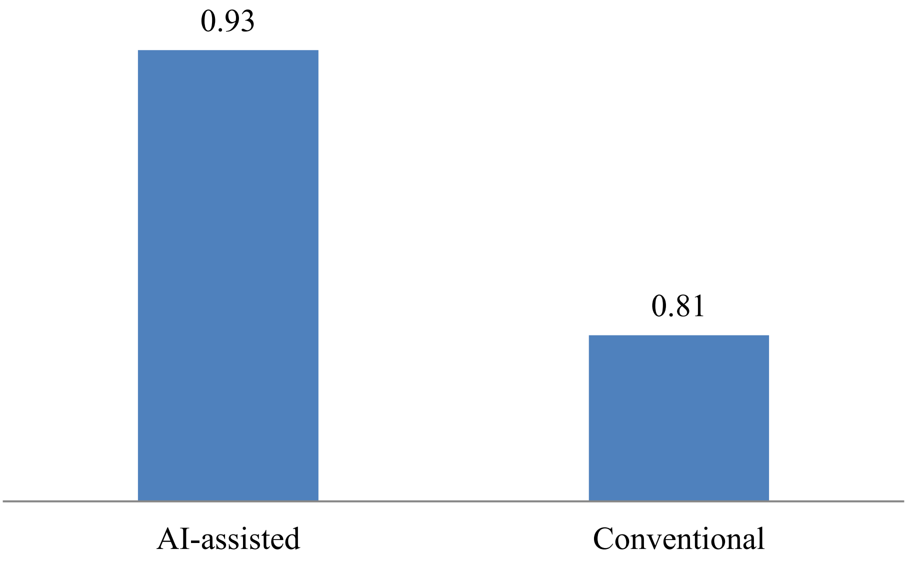

Figure 7: ROC comparing AI-assisted and conventional prediction of surgical difficulty

IV. Discussion

The present study assessed the role of an AI-based approach in estimating the surgical difficulty of impacted mandibular third molar removal and compared its performance with conventional clinical and radiographic assessment. The findings highlight the multifactorial nature of third molar surgery and support the potential value of predictive models capable of integrating multiple clinical and radiographic variables during preoperative evaluation.

Conventional assessment commonly relies on radiographic classification systems such as the Pederson and Pell–Gregory indices. Although these classifications are simple and widely used to describe the position and depth of impacted third molars, their ability to accurately predict operative difficulty remains limited. Akadiri et al. (2009) [10] reported that the Pederson index did not provide a sufficiently reliable prediction of third molar extraction difficulty. Similarly, García et al. (2000) [11] found that the Pell–Gregory classification was not a dependable predictor of the difficulty associated with impacted mandibular third molar removal. These findings suggest that tooth position alone may not adequately represent the technical challenges encountered during surgery.

Further evidence was provided by Diniz-Freitas et al. (2007) [12], who demonstrated that the Pederson scale was unable to accurately predict the difficulty of mandibular third molar extraction. Surgical complexity may instead reflect the combined influence of root morphology, extent of bone coverage, accessibility of the tooth, patient characteristics, and its relationship with surrounding anatomical structures. Therefore, assessment based on a limited number of predefined radiographic parameters may not fully represent the expected operative difficulty.

The multifactorial nature of third molar extraction was also emphasized by Susarla and Dodson (2004) [13], who identified several patient- and tooth-related factors associated with extraction difficulty. These findings provide a rationale for predictive approaches that can evaluate multiple variables simultaneously. In contrast to conventional scoring systems, machine-learning algorithms can process numerous predictors and identify complex relationships among them, potentially providing a more individualized estimate of surgical difficulty.

The application of AI to mandibular third molar surgery has recently received increasing attention. Khorshidi et al. (2024) [14] investigated the use of AI for predicting the difficulty of mandibular third molar extraction using CBCT-based information and demonstrated the potential of AI-based methods for preoperative assessment. Such approaches may assist in identifying combinations of clinical and radiographic characteristics associated with increased surgical difficulty and may provide additional information beyond conventional assessment methods.

Similarly, Yu D et al. (2023) [15] developed and validated machine-learning models for predicting postoperative pain following mandibular third molar surgery. Their use of machine-learning–based prediction is particularly relevant to clinical practice because identifying factors associated with postoperative risk may improve risk stratification and facilitate appropriate clinical interpretation. These findings support the potential role of machine-learning systems as decision-support tools in the preoperative assessment of third molar surgery.

An important potential advantage of AI-assisted prediction is greater consistency in preoperative assessment. Conventional estimation may vary according to clinical experience and individual interpretation of radiographic findings, particularly when multiple variables must be considered simultaneously. A suitably trained and validated AI model can apply a standardized analytical process across cases, potentially reducing subjective variation. However, AI-generated predictions should complement rather than replace professional clinical judgment, especially in cases with unusual anatomical or clinical characteristics.

From a clinical perspective, more accurate prediction of surgical difficulty may assist with treatment planning, patient counseling, selection of appropriate surgical expertise, anticipation of operative requirements, and preparation for cases with a greater likelihood of complications. Individualized risk estimation may also help clinicians determine when additional diagnostic evaluation or specialist management is appropriate.

Despite these potential benefits, implementation of AI requires careful validation. Model performance may be affected by differences in patient characteristics, imaging quality and protocols, surgical techniques, operator experience, and the criteria used to define surgical difficulty. Consequently, performance observed within a specific study population may not necessarily be reproduced in other clinical settings. External validation using larger and more diverse datasets is therefore necessary before widespread clinical implementation.

Overall, the findings of the present study indicate that conventional radiographic classification systems may have limited ability to fully predict the complexity of mandibular third molar surgery. The multifactorial nature of extraction difficulty suggests that assessment based solely on predefined radiographic parameters may not adequately reflect the clinical and anatomical factors involved. Recent advances in AI and machine-learning techniques have demonstrated promising potential for predicting surgical difficulties and postoperative complications. These approaches may complement conventional assessment by integrating multiple relevant variables and providing a more consistent and individualized estimation of surgical risk. Consequently, AI-assisted prediction could serve as a useful adjunct in preoperative planning and clinical decision-making for patients undergoing impacted mandibular third molar surgery.

V. Limitations

The present study has several limitations. First, the sample size was limited to 100 patients, which may reduce the precision and generalizability of the predictive estimates, particularly across individual radiographic subgroups. Second, the study was based on a single clinical dataset, and the findings may not be directly applicable to patients with different demographic characteristics, radiographic patterns, or treatment settings. Third, the performance of the AI-assisted method was evaluated against the observed surgical difficulty within the study cohort; therefore, external validation using independent datasets is necessary to determine its reproducibility and clinical robustness. In addition, variations in surgical expertise, case selection, and assessment of operative difficulty may have influenced the observed outcomes. Finally, although the AI-assisted approach demonstrated superior predictive performance, these results should not be interpreted as evidence that AI can replace clinical judgment. Prospective multicenter studies with larger and more diverse populations, external validation, and assessment of clinical outcomes are warranted before routine clinical implementation.

VI. Conclusion

The present study demonstrates that the AI-assisted method provided more accurate and consistent prediction of surgical difficulty in impacted mandibular third molar removal than the conventional approach. By enhancing preoperative assessment of surgical complexity, AI-assisted evaluation may serve as a valuable adjunct to conventional radiographic assessment and support more informed treatment planning. However, validation in larger, multicenter, and independent patient populations is warranted before routine clinical implementation**.**

Conflicts of Interest: Nil
Financial Support: None

References

[1]. Susarla SM, Dodson TB. Risk factors for third molar extraction difficulty. J Oral Maxillofac Surg. 2004 Nov;62(11):1363-71. doi: 10.1016/j.joms.2004.05.214. PMID: 15510357.

[2]. Renton T, Smeeton N, McGurk M. Factors predictive of difficulty of mandibular third molar surgery. Br Dent J. 2001 Jun 9;190(11):607-10. doi: 10.1038/sj.bdj.4801052. PMID: 11441899.

[3]. Diniz-Freitas M, Lago-Méndez L, Gude-Sampedro F, Somoza-Martin JM, Gándara-Rey JM, García-García A. Pederson scale fails to predict how difficult it will be to extract lower third molars. Br J Oral Maxillofac Surg. 2007 Jan;45(1):23-6. doi: 10.1016/j.bjoms.2005.12.004. Epub 2006 Jan 24. PMID: 16434132.

[4]. Bali A, Bali D, Sharma A, Verma G. Is Pederson Index a True Predictive Difficulty Index for Impacted Mandibular Third Molar Surgery? A Meta-analysis. J Maxillofac Oral Surg. 2013 Sep;12(3):359-64. doi: 10.1007/s12663-012-0435-x. Epub 2012 Sep 22. PMID: 24431870; PMCID: PMC3777040.

[5]. Susarla SM, Dodson TB. How well do clinicians estimate third molar extraction difficulty? J Oral Maxillofac Surg. 2005 Feb;63(2):191-9. doi: 10.1016/j.joms.2004.05.220. PMID: 15690287.
[6]. Sánchez-Torres A, Soler-Capdevila J, Ustrell-Barral M, Gay-Escoda C. Patient, radiological, and operative factors associated with surgical difficulty in the extraction of third molars: a systematic review. Int J Oral Maxillofac Surg. 2020 May;49(5):655-665. doi: 10.1016/j.ijom.2019.10.009. Epub 2019 Nov 15. PMID: 31735527.
[7]. Qiu P, Huang J, Zhang H, Liu S, Dong B, Zhang X. A machine learning-based predictive model for mandibular third molar extraction difficulty: incorporating multimodal features and SHAP analysis. BMC Oral Health. 2026 Jan 19;26(1):309. doi: 10.1186/s12903-026-07701-3. PMID: 41555297; PMCID: PMC12903583.
[8]. Yoo JH, Yeom HG, Shin W, Yun JP, Lee JH, Jeong SH, Lim HJ, Lee J, Kim BC. Deep learning based prediction of extraction difficulty for mandibular third molars. Sci Rep. 2021 Jan 21;11(1):1954. doi: 10.1038/s41598-021-81449-4. PMID: 33479379; PMCID: PMC7820274.
[9]. Chindanuruks T, Jindanil T, Cumpim C, Sinpitaksakul P, Arunjaroensuk S, Mattheos N, Pimkhaokham A. Development and validation of a deep learning algorithm for the classification of the level of surgical difficulty in impacted mandibular third molar surgery. Int J Oral Maxillofac Surg. 2025 May;54(5):452-460. doi: 10.1016/j.ijom.2024.11.008. Epub 2024 Dec 4. PMID: 39632213.
[10]. Akadiri OA, Fasola AO, Arotiba JT. Evaluation of Pederson index as an instrument for predicting difficulty of third molar surgical extraction. Niger Postgrad Med J. 2009 Jun;16(2):105-8. PMID: 19606189.
[11]. García AG, Sampedro FG, Rey JG, Vila PG, Martin MS. Pell-Gregory classification is unreliable as a predictor of difficulty in extracting impacted lower third molars. Br J Oral Maxillofac Surg. 2000 Dec;38(6):585-587. doi: 10.1054/bjom.2000.0535. PMID: 11092770.
[12]. Diniz-Freitas M, Lago-Méndez L, Gude-Sampedro F, Somoza-Martin JM, Gándara-Rey JM, García-García A. Pederson scale fails to predict how difficult it will be to extract lower third molars. Br J Oral Maxillofac Surg. 2007 Jan;45(1):23-6. doi: 10.1016/j.bjoms.2005.12.004. Epub 2006 Jan 24. PMID: 16434132.
[13]. Susarla SM, Dodson TB. Risk factors for third molar extraction difficulty. J Oral Maxillofac Surg. 2004 Nov;62(11):1363-71. doi: 10.1016/j.joms.2004.05.214. PMID: 15510357.
[14]. Khorshidi F, Esmaeilyfard R, Paknahad M. Enhancing predictive analytics in mandibular third molar extraction using artificial intelligence: A CBCT-Based study. Saudi Dent J. 2024 Dec;36(12):1582-1587. doi: 10.1016/j.sdentj.2024.11.007. Epub 2024 Nov 26. PMID: 40952877; PMCID: PMC11976095.
[15]. Yu D, Liu Z, Zhuang W, Li K, Lu Y. Development and validation of machine learning based prediction model for postoperative pain risk after extraction of impacted mandibular third molars. Heliyon. 2023 Nov 30;9(12):e23052. doi: 10.1016/j.heliyon.2023.e23052. Erratum in: Heliyon. 2024 Nov 28;11(1):e40609. doi: 10.1016/j.heliyon.2024.e40609. PMID: 38076075; PMCID: PMC10703859.