Journal of Dental Sciences

Volume 21 | Issue 2 Article 8

2026

Artificial intelligence-enabled automatic segmentation of impacted mandibular third molars: A comprehensive comparison of multiple algorithms

Sum Yin Au Yeung

Wai Ying Kot

Yin Yan Leung

Ka Ho Chan

Wei-fa Yang

Follow this and additional works at: https://jds.ads.org.tw/journal

Recommended Citation
Yeung, Sum Yin Au; Kot, Wai Ying; Leung, Yin Yan; Chan, Ka Ho; and Yang, Wei-fa (2026) "Artificial intelligence-enabled automatic segmentation of impacted mandibular third molars: A comprehensive comparison of multiple algorithms," *Journal of Dental Sciences*: Vol. 21: Iss. 2, Article 8.
Available at: https://jds.ads.org.tw/journal/vol21/iss2/8

This Original Article is brought to you for free and open access by Journal of Dental Sciences. It has been accepted for inclusion in Journal of Dental Sciences by an authorized editor of Journal of Dental Sciences. For more information, please contact cpchiang@ntu.edu.tw.

Available online at https://jds.ads.org.tw/journal/

Digital Commons

journal homepage: https://jds.ads.org.tw/journal/

Original Article

Artificial intelligence-enabled automatic segmentation of impacted mandibular third molars: A comprehensive comparison of multiple algorithms

Sum Yin Au Yeung, Wai Ying Kot, Yin Yan Leung, Ka Ho Chan, Pui Hang Leung, Wei-fa Yang*

Faculty of Dentistry, The University of Hong Kong, 34 Hospital Road, 999077, Hong Kong SAR, China

Received 5 October 2025; Final revision received 27 October 2025
Available online 1 April 2026

KEYWORDS
Artificial intelligence;
Cone-beam computed tomography;
Mandibular third molars;
Tooth segmentation;
Wisdom tooth

Abstract *Background/purpose:* Impacted mandibular third molar models generated from cone-beam computed tomography images facilitates preoperative assessment and surgical planning. Despite recent applications of artificial intelligence (AI) in tooth segmentation, its performance in mandibular third molar segmentation remains underexplored. This study aims to develop a comprehensive comparison of available algorithms for mandibular third molar segmentation for clinical applications and research.

Materials and methods: Forty impacted mandibular third molars were segmented using interactive thresholding (Thresholding), Blue Sky Plan (BSP AI), Blue Sky Plan with manual adjustments (BSP AI with adjustment), and Planmeca Romexis (Romexis AI), with manual segmentation as reference standard. Segmentation accuracy was assessed using dice similarity coefficient (DSC), intersection over union (IoU), average symmetric surface distance (ASSD), 95 % Hausdorff distance (95HD) and relative volume difference (RVD). Segmentation quality was evaluated based on morphology, three-dimensional visualization, and spatial relationships with adjacent structures.

Results: BSP AI with adjustment achieved the highest accuracy (DSC: 0.89 ± 0.02, 95HD: 0.85 ± 0.25 mm), followed by Thresholding (DSC: 0.86 ± 0.03; 95HD: 1.28 ± 0.23 mm) and BSP AI (DSC: 0.85 ± 0.08; 95HD: 1.91 ± 1.91 mm). Romexis AI (DSC: 0.81 ± 0.03; 95HD: 1.41 ± 0.32 mm) demonstrated the poorest performance. BSP AI with adjustment achieved the highest quality score of 38.00 (38.00–39.00), similar to Thresholding achieving 38.00 (37.25–39.00), followed by BSP AI and Romexis AI being the last.

Conclusion: AI-enabled segmentation achieved high accuracy and efficiency in mandibular third molar segmentation, while manual adjustment ensured optimal quality. Further research

* Corresponding author. Faculty of Dentistry, The University of Hong Kong, 34 Hospital Road, Hong Kong SAR 999077, China.
E-mail address: teddyrun@hku.hk (W. Yang).

https://doi.org/10.1016/j.jds.2025.10.032
1991-7902/© 2026 Association for Dental Sciences of the Republic of China. Publishing services by Digital Commons. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

is warranted to promote future enhancement and clinical applications of AI-enabled mandibular third molar segmentation.
© 2026 Association for Dental Sciences of the Republic of China. Publishing services by Digital Commons. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Surgical removal of mandibular third molars is one of the most frequently performed procedures in oral and maxillofacial surgery,[1] which is frequently associated with a risk of damaging adjacent vital structures, such as the inferior alveolar nerve and the adjacent second molar.[2,3] While panoramic radiographs have been recommended for the preoperative evaluation,[1,4] its two-dimensional (2D) nature often makes it challenging to visualise the spatial relationship when the third molar overlaps adjacent structures.[5] In such cases, cone-beam computed tomography (CBCT) provides three-dimensional (3D) examination of the mandibular third molar and adjacent vital structures.[5,6]

CBCT images are traditionally viewed in 2D slices, requiring high level of spatial visualisation skills for 3D reconstruction mentally. Although there is lack of robust evidence supporting the use of CBCT in reducing the risk of damaging adjacent structures,[7,8] the innovative 3D modelling from CBCT segmentation provides a valuable tool for direct visualisation, enhancing comprehension of the spatial relationship and aiding preoperative risk assessment. These 3D models can be applied in virtual surgical planning to determine surgical strategies,[9] facilitating dynamic navigation and improving intraoperative efficiency.[10-12] Moreover, 3D printing can produce customised 3D models and surgical guides for simulation, enhancing surgeon confidence.[13–15] Given the recent emergence of this approach, further clinical research is necessary to confirm its benefits in preoperative planning.

Tooth segmentation has been conventionally performed manually by outlining the structure from CBCT images slices. This approach is time-consuming and highly dependent on the operator experience.[16] Consequently, semi-automatic segmentation techniques were developed to reduce operator input and time required,[17] with interactive thresholding being one of the most widely utilised method. This method separates the target object by defining a specific radiodensity range, which is effective for bone segmentation due to the radiodensity difference between bone and soft tissues.[18] However, it is less popular in tooth segmentation due to the comparable radiodensities between cementum and alveolar bone, making differentiation challenging.[19]

Concurrently, there has been an increased trend in the integration of artificial intelligence (AI) into the segmentation workflow. In our systematic review, we have found that deep learning algorithms, mainly convolutional models with U-Net architecture, achieved high accuracies in tooth segmentations with their capacity in manipulating fine details.[20] In addition, there has been an increasing application of Transformer-based algorithms for their superior ability to identify global relationships, overcoming the limitations of U-Net architectures.[20] Recent applications of internally developed deep learning algorithms in mandibular third molars segmentation achieving dice similarity coefficients (DSC) of over 90 %.[21–23] However, there is a lack of research comparing the segmentation accuracy of these deep learning algorithms against traditional methods specifically for impacted mandibular third molars in clinical settings. This study aims to provide a comparative evaluation and clinical benchmarking of existing available segmentation methods applicable to impacted mandibular third molar segmentation, with a focus in the segmentation accuracy, quality and efficiency.

Materials and methods

This study evaluated four clinically available segmentation methods for mandibular third molars based on their segmentation accuracy, quality and efficiency. The study workflow is illustrated in Fig. 1.

Case selection

A total of 20 CBCT scans, including 40 mandibular third molars, were obtained from the database of The Prince Philip Dental Hospital. The scans were taken prior to mandibular third molar surgery using the Planmeca ProMax 3D Mid (Planmeca, Helsinki, Finland) machine with settings of: 0.4 mm voxel size, field of view of 20 cm diameter and 10 cm height, and 229 mGy*cm^{2}dose area product. Inclusion criteria included: (1) impacted mandibular third molars; (2) existing CBCT scans capturing both mandibular third molars; (3) optimal quality without obvious artifacts or distortions; and (4) adjacent mandibular second molars present. Scans with cystic lesions involved or from vulnerable patients were excluded. The selected CBCT scans were anonymised by removing all patient details and identifiers, and then exported in DICOM file format for segmentation. Ethics approval was obtained from The Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (HKU/HA HKW IRB) (Reference Number: UW 23—466). Written consent has been waived.

Tooth segmentation

The reference standard was established by manual segmentation with 3D Slicer (Harvard Medical School, Boston, MA, USA; Version 5.6.2), an open-source software.[24] Reference models were generated by outlining the tooth on

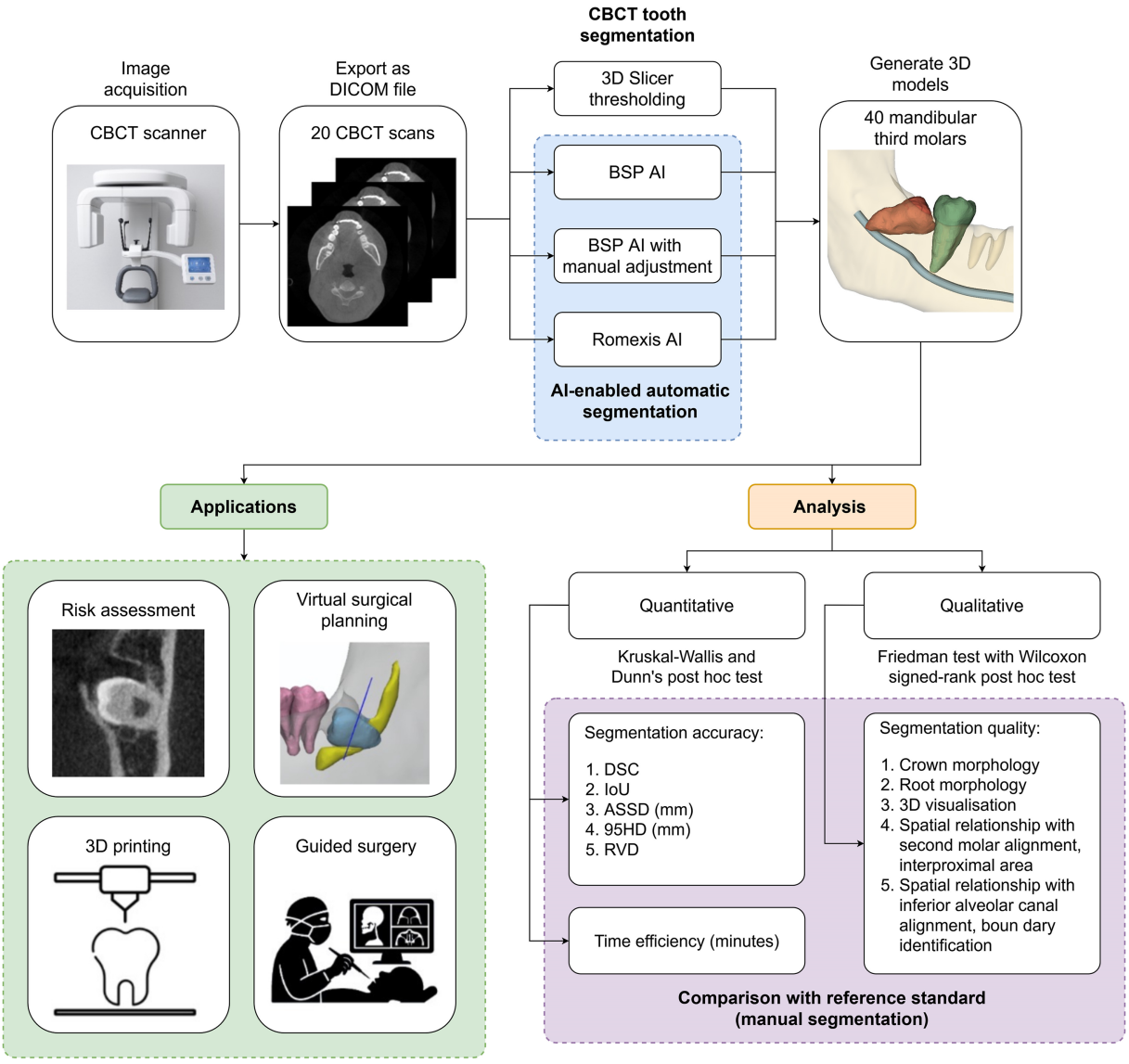

Figure 1 Illustration of the study flow and applications of mandibular third molar segmentation.
Note: CBCT: Cone-beam computed tomography; DSC: Dice similarity coefficient; IoU: Intersection over union; ASSD: Average symmetric surface distance; 95HD: 95 % Hausdorff distance; RVD: Relative volume difference.

all CBCT slices manually, then verified by a second operator with reference to the original CBCT image. Four clinically available segmentation methods were compared in this study, as outlined in Table 1. Interactive thresholding (Thresholding) was conducted with 3D Slicer,[24] by selecting a threshold interval and manually removing unwanted regions (Supplementary Fig. 1). AI-enabled automatic segmentation was performed with Blue Sky Plan (Blue Sky Bio, Libertyville, IL, USA; Version 4.13.34) (BSP AI). Subsequent manual adjustments were performed on the generated segmentation (BSP AI with adjustment) for comparison (Supplementary Fig. 2). Another AI-enabled segmentation method included, Planmeca Romexis (Planmeca, Helsinki, Finland; Version 6.1) (Romexis AI), involved the operator annotating the tooth to guide the AI-enabled automatic fill to complete the segmentation (Supplementary Fig. 3). The segmentations were then transferred to 3D Slicer for comparison.[24]

Segmentation accuracy

To compare the accuracy of segmentations generated with the four methods objectively, five evaluation metrics were used. Evaluation metrics can be classified into three main groups: overlap-based, distance-based, and volume-based metrics. Overlap-based metrics measure the spatial overlap between the segmentation model and the reference standard. The dice similarity coefficient (DSC) and the intersection over union (IoU) are the two most commonly used overlap-based metrics, both ranging from zero to one, with higher values indicating greater agreement. DSC calculates the percentage of overlap between the sum of the compared segmentation and reference standard:[25]

$$DSC = \frac{2TP}{2TP + FP + FN}$$

Table 1 Characteristics of the segmentation software and methods compared in this study.

Software	Provider and Version	Segmentation methods
3D Slicer	Surgical Planning Lab, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA (Version 5.6.2)	**Reference standard:** Manual segmentation by outlining the tooth on all CBCT images slices **Thresholding:** Interactive thresholding followed by removal of unwanted parts manually
Blue Sky Plan	Blue Sky Bio, Libertyville, IL, USA (Version 4.13.34)	**BSP AI:** AI-enabled automatic segmentation by clicking any point on the tooth from any view to select the tooth to be segmented **BSP AI with adjustment:** Manual adjustment with refinement tools after AI-enabled automatic segmentation
Romexis	Planmeca, Helsinki, Finland (Version 6.1)	**Romexis AI:** AI-enabled automatic segmentation by masking the selected tooth with tooth segmentation tools

Similarly, IoU, otherwise referred to as Jaccard Index, quantifies the proportion of intersection divided by union:[25]

$$IoU = \frac{TP}{TP + FP + FN}$$

Distance-based metrics measures the spatial distance between the model surfaces in millimetres, with lower values reflecting closer alignment. The average symmetric surface distance (ASSD) and the 95 % Hausdorff distance (95HD) values were obtained with the ModelToModelDistance and SlicerRT extension in 3D Slicer respectively.[26]

Lastly, for volume-based metrics, relative volume difference (RVD) was used to determine the percentage discrepancy between the size of the segmentation model and reference standard. RVD values are calculated using the following formula, ranging range from negative one zero to one:

$$RVD = \frac{V_S - V_R}{V_R}$$

where V_S and V_R represents the volume of the segmentation model and reference standard respectively.

Segmentation quality

To evaluate the visual and spatial features of the segmentation models, an assessment rubric was developed with a focus on segmentation quality. The scoring table (Table 2) included the evaluation of the visual quality of the crown and root morphologies and their spatial relationship with adjacent structures, inferior alveolar canal and mandibular second molar. Each aspect was assigned a score from 1 to 4, with 4 representing excellent quality. The overall score is achieved by adding up the scores, indicating the overall qualitative performance and its suitability for clinical purposes. The models were evaluated by two assessors individually, with a third assessor verifying the scores and resolving any disagreements.

Time spent

Besides evaluation metrics, the time required for segmentation was recorded in minutes, excluding the time needed for importing CBCT images and exporting the final model.

Statistical analysis

Segmentation accuracy was obtained by calculating the means, standard deviations, medians and interquartile ranges of the five evaluation metrics. The Shapiro–Wilk Normality Test was performed to assess the normal distribution of data. In view of the non-normal distribution indicated by the test, Kruskal–Wallis test was conducted to compare the four methods, followed by Dunn's post hoc tests with Holm correction to control for multiple testing. For quality scores, medians and interquartile ranges were obtained. Since the Shapiro–Wilk Normality Test indicated that the data were non-normal, Friedman test was conducted for comparison, with Wilcoxon signed-rank post hoc tests for pairwise comparisons. Statistical significance was set at $P < 0.05$ for all analyses. Statistical analyses were performed using IBM SPSS Statistics for Windows (IBM Corp., Armonk, NY, USA; Version 29.0.2.0).

Results

The segmentation accuracy of the four methods are compared in Table 3. Overall, BSP AI with adjustment achieved the best performance across all metrics, achieving the highest overlap values (DSC: 0.89 ± 0.02; IoU 0.81 ± 0.30), which were statistically significantly higher than all other methods ($P < 0.05$). This method also yielded the shortest boundary distance discrepancies (ASSD: 0.32 ± 0.07 mm; 95HD: 0.85 ± 0.25 mm), comparable to Thresholding. While BSP AI and Thresholding showed statistically comparable results for overlap metrics, Romexis AI consistently yielded the lowest accuracy. RVD values showed that both BSP AI and BSP AI with adjustment introduced the least systematic volume bias.

Fig. 2 illustrates a case comparing mandibular third molars models generated by the four methods with the reference standard. Based on the quality scores (Supplementary Table 1), BSP AI with adjustment yielded

Table 2 Scoring rubrics for assessing the quality of the segmentation models.

Criteria	Poor (1)	Fair (2)	Good (3)	Excellent (4)
1. Crown morphology				
Crown shape	Pronounced distortions or irregularities in shape.	Noticeable deviations compromising anatomical accuracy.	Slight deviations with overall form intact.	Accurate representation aligning with anatomical norms.
Crown size	Major discrepancies, grossly disproportionate.	Moderate alterations detracting from proportionality.	Minor discrepancies not greatly impacting representation.	Precise and proportional with excellent accuracy.
Cusps & grooves	Severe loss of detail or distortion.	Partial loss of detail with reduced clarity.	Generally well-defined with minimal imperfections.	Highly detailed and anatomically precise.
2. Root morphology				
Number of roots	Incorrect identification with severe distortion.	Visible root structure with noticeable inaccuracies.	Correct identification with minor loss of detail.	Precise identification of roots with no inaccuracies.
Root anatomy	Poorly defined boundaries with major morphological features missing and grossly incorrect	Visible root structure with absence or distortion of finer details.	Generally accurate identification of curvatures and apices with minor defects.	Highly detailed identification of root anatomy with detailed curvatures and root apices.
3.3D Visualisation				
3D positioning	Grossly misaligned with significant errors.	Moderate inaccuracies with noticeable deviations.	Minor inaccuracies with slight deviations.	Precise accuracy and alignment.
4. Spatial relationship with mandibular second molar				
Alignment	Grossly misaligned with significant errors.	Moderate inaccuracies with noticeable deviations.	Minor inaccuracies with slight deviations.	Precise accuracy and alignment.
Interproximal area	Severely distorted or unrecognizable.	Partially defined with irregular boundaries.	Mostly well-defined with minor inaccuracies.	Precisely defined with accurate integration.
5. Spatial relationship with inferior alveolar canal				
Alignment	Grossly misaligned with significant errors.	Moderate inaccuracies with noticeable deviations.	Minor inaccuracies with slight deviations.	Precise accuracy and alignment.
Boundary identification	Severely distorted or unrecognizable.	Partially defined with irregular boundaries.	Mostly well-defined with minor inaccuracies.	Precisely defined with accurate integration.

Note.
Each section is evaluated based on the provided criteria, and a score is assigned to each criterion. The scores within each section are summed up for a section score. An overall assessment score (maximum 40) is obtained by adding up the scores of each section. A score of 35–40 implies excellent execution of segmentation, highly suitable for clinical evaluations and research. A score range of 28–34 indicates good segmentation, meeting most clinical and research requirements. A score of 20–27 implies that the segmentation has deviations which limits its reliability in clinical and research applications. A score of 19 or below reflects poor segmentation quality, which is deemed unsuitable for reliable interpretations.

the highest overall 38.32 ± 0.87 out of 40, producing accurate anatomical features and precise spatial alignment. Thresholding scored the second highest (38.00 ± 1.14), with slight distortions in fine details at apical region, followed by BSP AI (35.87 ± 3.21). Romexis AI had the lowest score (31.45 ± 1.35) with less detailed morphology and poorly defined root anatomies. While all methods obtained high scores in visualisation and spatial relationship, as shown in Fig. 3, their performance varied in the identification of tooth anatomies, with statistical significance found in aspects evaluating crown morphology, root morphology, and boundary identification (Supplementary Table 1).

In terms of the time spent, manual segmentation was the most time-consuming, requiring 43.78 ± 4.76 min per tooth. Among the four tested methods, BSP AI required the least time, with a mean duration of only 1 min. The time

Table 3 Evaluation of segmentation accuracy of the four segmentation methods.

	Thresholding		BSP AI		BSP AI with adjustment		Romexis AI	
	Mean ± SD	Median (IQR)	Mean ± SD	Median (IQR)	Mean ± SD	Median (IQR)	Mean ± SD	Median (IQR)
DSC (0–1)	0.86 ± 0.03 [b]	0.86 (0.85 −0.87)	0.85 ± 0.08 [b]	0.88 (0.85 −0.89)	0.89 ± 0.02[a]	0.89 (0.88 −0.91)	0.81 ± 0.03[c]	0.82 (0.80 −0.83)
IoU (0–1)	0.75 ± 0.05 [b]	0.75 (0.73 −0.77)	0.74 ± 0.11 [b]	0.78 (0.74 −0.80)	0.81 ± 0.30[a]	0.81 (0.79 −0.83)	0.69 ± 0.06[c]	0.69 (0.67 −0.71)
ASSD (mm)	0.39 ± 0.08 [b]	0.39 (0.34 −0.44)	0.59 ± 0.57 [b]	0.41 (0.35 −0.46)	0.32 ± 0.07[a]	0.32 (0.27 −0.36)	0.58 ± 0.10[c]	0.57 (0.52 −0.62)
95HD (mm)	1.28 ± 0.23 [b]	1.27 (1.13 −1.39)	1.91 ± 1.91 [b]	1.24 (1.02 −1.64)	0.85 ± 0.25[a]	0.79 (0.66 −0.94)	1.41 ± 0.32 [b]	1.32 (1.22 −1.55)
RVD ((-1)-1)	−0.13 ± 0.08 [b]	−0.13 ((-0.18)-(-0.09))	0.06 ± 0.32[a]	0.00 ((-0.10)-0.06)	−0.01 ± 0.07[a]	−0.01 ((-0.05)-0.04)	−0.13 ± 0.08 [b]	−0.14 ((-0.18)-(-0.07))

Note.
Differences between groups are statistically significant ($P < 0.05$) for all evaluation metrics according to Kruskal–Wallis tests. a, b, c are identifiers for post hoc comparison tests within each evaluation metric, and means with different superscripts differ significantly at $P < 0.05$ in Dunn's post hoc test. DSC: Dice similarity coefficient; IoU: Intersection of union; ASSD: Average symmetric surface distance; 95HD: 95 % Hausdorff distance; RVD: Relative volume difference.

spent for segmentation using Romexis AI was 5.50 ± 1.18 min, Thresholding was 8.89 ± 1.68 min and BSP AI with adjustment was 14.75 ± 3.51 min per tooth.

Discussion

In this study, a comprehensive assessment protocol was developed to compare the performance of four different segmentation algorithms in generating 3D models of mandibular third molars from CBCT images. Evaluation criteria included segmentation accuracy, quality and efficiency. BSP AI with adjustment outperformed the other three methods in both quantitative and qualitative aspects.

Conventionally, manual segmentation was the adopted method of segmentation, which is highly reliant on operator experience to accurately outline the tooth from each CBCT slice. This process can be time-consuming and tedious.[16] To enhance efficiency, various semi-automatic algorithms were developed, with Thresholding being one of the most widely used methods. It is effective for bone segmentation due to high radiodensity differences between bone and soft tissues.[18] However, its application in tooth segmentation is limited by the similar densities between cementum and alveolar bone, making it challenging to separate the tooth from surrounding structures.[19,27] Despite this, Thresholding achieved results comparable to BSP AI with adjustment in our study, highlighting the importance of operator input in achieving high-quality segmentation, with manual verification resulting in increased accuracy in low-contrast areas. Both methods were significantly more efficient than traditional manual segmentation, suggesting that Thresholding can also be highly efficient in segmentation.

In recent years, the integration of AI in segmentation has markedly increased, aiming to achieve both accuracy and efficiency. Among deep learning algorithms, U-Net architecture and its derivatives are the most popular algorithms due to their ability in handling complex data and fine details.[20] Despite their extensive applications, their performance in mandibular third molars segmentation remains relatively unknown due to limited studies available.

Two AI-enabled segmentation software were compared in this study. Planmeca Romexis (Planmeca), a popular CBCT software included with Planmeca dental CBCT units, utilises a deep learning algorithm employing convolutional neural networks (CNNs) trained and validated with panoramic radiograph datasets.[28] It also supports standard DICOM data from other CBCT machines, enhancing its clinical utility. The second software, Blue Sky Plan (Blue Sky Bio), provides complimentary planning and surgical guide design tools, and has gained popularity with its user-friendly interface and powerful tools.[29] Currently, there are no details regarding the algorithm adopted. Noting that it offers segmentation refinement tools, BSP AI with adjustment was included to evaluate segmentation accuracy with minimal operator input.

Comparing the two AI-enabled segmentation methods, BSP AI achieved statistically significantly higher accuracy than Romexis AI in overlap-based metrics. Given the observed effect size of Cohen's $d = 0.54$, it is important that other factors, such as the visual quality, should be considered when evaluating the clinical significance of the

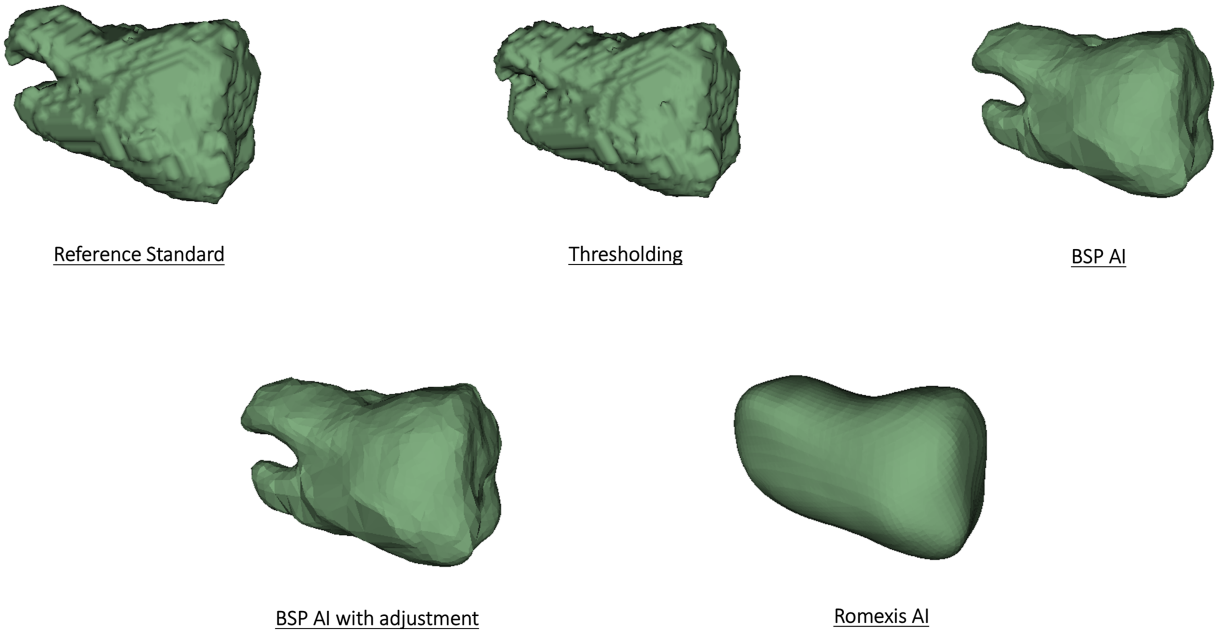

Figure 2 Demonstration of the segmentation models generated by the four segmentation methods in a random case.

models. Despite the currently lack of knowledge regarding a clinically acceptable DSC value, quality scores have shown that Romexis AI is inadequate at identifying root morphology, affecting the interpretation of apical regions. Given that Romexis AI obtained a DSC value of 81 %, it is proposed that the clinically acceptable DSC value should be between 81 % and 85 %, with further careful evaluation of the models required for clinical applications. While Romexis AI uses CNNs, previous research showed that CNNs may achieve lower segmentation accuracy than other algorithms such as U-Net and Transformer-based architectures due to their limited effectiveness in training complex data.[20] However, both methods are still inadequate in identifying fine details at segmentation boundaries, as demonstrated by distance-based metrics, which is essential in assessing its relationship with adjacent structures.

In this study, DSC values of AI-enabled automatic segmentations ranged from 0.81 to 0.89, which is slightly lower than 0.920 to 0.983 reported in recent studies on automatic mandibular third molar segmentation.[21–23] This may be explained by the variations of segmentation algorithms. The majority adopted deep learning algorithms which were internally trained and optimised for the differentiation and segmentation of mandibular third molars until it reaches the target performance.[30,31] However, these algorithms lack external validation, which is crucial for evaluating performance in different clinical settings. In contrast, methods included in this study are designed for segmentation of all tooth types. Nevertheless, considering that a DSC value between 0.8 and 0.9 is accepted as excellent agreement,[32] our results showed a high percentage of overlap with the reference standard.

In recent studies, qualitative assessments have been used together with quantitative assessments to evaluate segmentations of the maxillofacial region.[33–35] However, these scoring systems only give an overall indication of their clinical applicability. To address the limitations, we have developed a detailed evaluation protocol focusing on morphology and spatial relationship with adjacent structures, providing a more comprehensive assessment to facilitate surgical planning. Models generated with BSP AI with adjustment and Thresholding showed well-defined anatomy, while those by BSP AI and Romexis AI had more distortions at apical region. Some BSP AI models showed severely distorted anatomies, including segmentation of the opposing teeth or missing roots, while Romexis AI models tend to be underestimated and excessively smoothened, resulting in difficulties in identifying anatomical features and subsequent determination of its relationship with adjacent structures. Despite varying tooth model quality, the performance in identifying its relationship with adjacent structures remained consistent, likely due to a relatively high contrast between mandibular third molars and adjacent structures.

Considering both quantitative and qualitative aspects, this study presents a comprehensive assessment of the clinical applicability of mandibular third molar segmentation methods. In clinical scenarios, these customised models can provide 3D visualisation of the surgical region for virtual surgical planning to identify possible risks such as root fracture and damage to adjacent structures, thereby improving surgical efficiency and outcomes.[9] Additionally, the models can be 3D printed for simulation to replicate the surgical procedure in complex cases.[13] It can also be presented to patients prior to surgery for better understanding.

A potential limitation of this study is the possible intra-subject correlation, in which the 40 mandibular third molars originated from 20 patients with bilateral mandibular third molars. Despite this, it is hypothesised that the accuracy of the samples will not be significantly affected, given that the CBCT scans were generated from the same

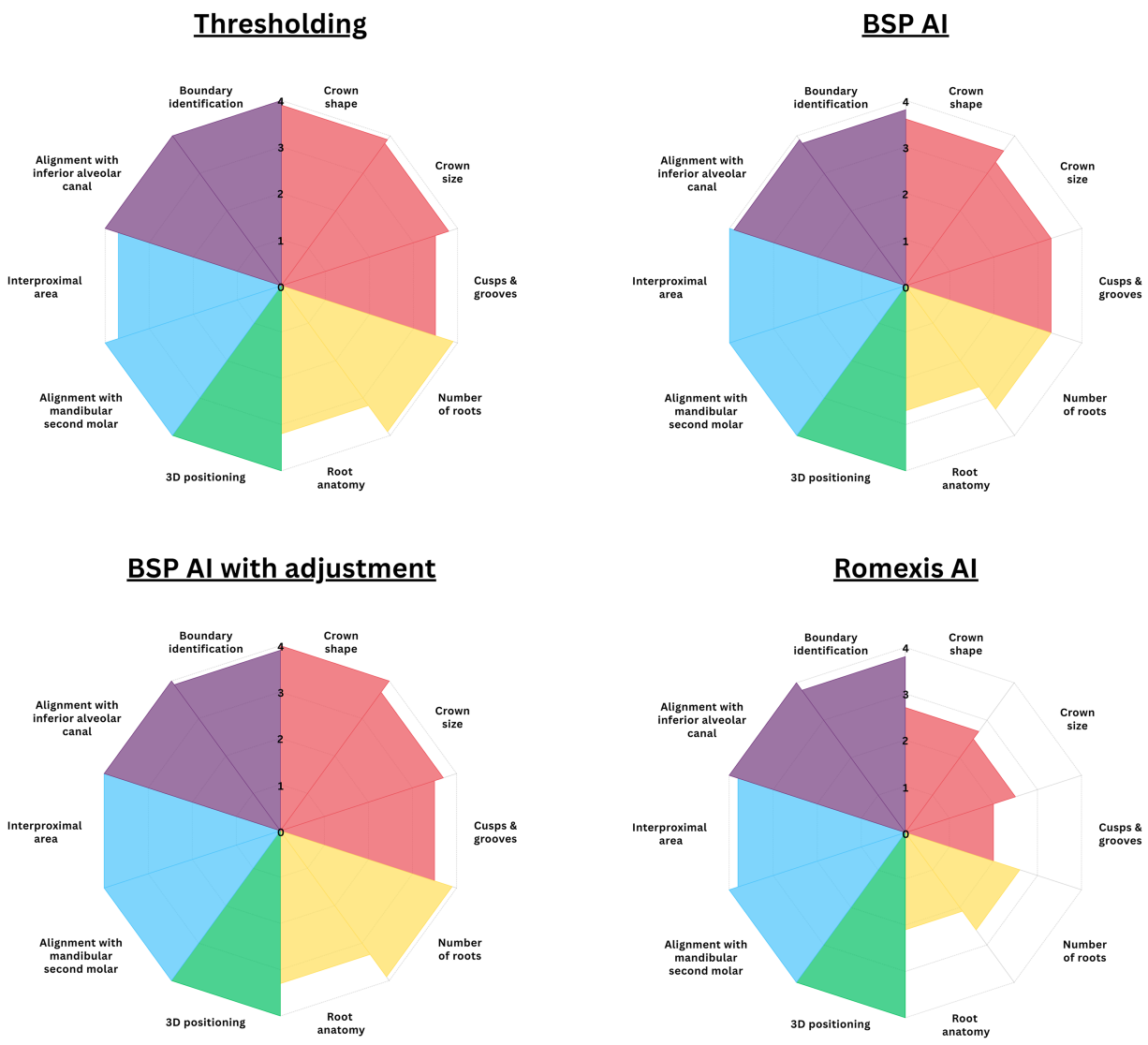

Figure 3 Comparison of the segmentation quality of the four segmentation methods with reference to manual segmentation.

machine with the same settings, with mandibular third molars having distinct morphology and positioned on opposite sides of the dental arch without any overlapping regions in between. Moreover, the majority of the estimated Intraclass Correlation Coefficient (ICC) values were not significantly different from zero, indicating negligible patient-level clustering effects. Another limitation is the relatively small sample size of 40 mandibular third molars. Nevertheless, since AI-enabled automatic segmentation software are tested and validated with an internal dataset before release, the 40 samples serve as an independent external dataset for evaluating the segmentation accuracy. Moreover, clinical aspects have been evaluated visually and spatially, exploring their potential for future clinical application. In view of the rapid development of both open-source and commercially available AI-enabled segmentation software, we propose that future research should focus on exploring the integration of mandibular third molar segmentations with these software under clinical settings, employing comprehensive evaluation metrics of segmentation accuracy, quality and efficiency.

Declaration of competing interest

The authors have no conflicts of interest relevant to this article.

Acknowledgments

This study is partially supported by Health and Medical Research Funds (11221966, 12232366), Food and Health Bureau, Hong Kong.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jds.2025.10.032.

References

1. Leung YY, Hung KF, Li DTS, Yeung AWK. Application of cone beam computed tomography in risk assessment of lower third molar surgery. *Diagnostics* 2023;13:919.
2. Cheung LK, Leung YY, Chow LK, Wong MC, Chan EK, Fok YH. Incidence of neurosensory deficits and recovery after lower third molar surgery: a prospective clinical study of 4338 cases. *Int J Oral Maxillofac Surg* 2010;39:320–6.
3. McArdle LW, Patel N, Jones J, McDonald F. The mesially impacted mandibular third molar: the incidence and consequences of distal cervical caries in the mandibular second molar. *Surgeon* 2018;16:67–73.
4. Ghaeminia H, Meijer GJ, Soehardi A, Borstlap WA, Mulder J, Bergé SJ. Position of the impacted third molar in relation to the mandibular canal. Diagnostic accuracy of cone beam computed tomography compared with panoramic radiography. *Int J Oral Maxillofac Surg* 2009;38:964–71.
5. Patel PS, Shah JS, Dudhia BB, Butala PB, Jani YV, Macwan RS. Comparison of panoramic radiograph and cone beam computed tomography findings for impacted mandibular third molar root and inferior alveolar nerve canal relation. *Indian J Dent Res* 2020;31:91–102.
6. Reia VCB, de Toledo Telles-Araujo G, Peralta-Mamani M, Biancardi MR, Rubira CMF, Rubira-Bullen IRF. Diagnostic accuracy of CBCT compared to panoramic radiography in predicting IAN exposure: a systematic review and meta-analysis. *Clin Oral Invest* 2021;25:4721–33.
7. de Toledo Telles-Araújo G, Peralta-Mamani M, Caminha RDG, et al. CBCT does not reduce neurosensory disturbances after third molar removal compared to panoramic radiography: a systematic review and meta-analysis. *Clin Oral Invest* 2020;24:1137–49.
8. Klatt JC, Sorowka T, Kluwe L, Smeets R, Gosau M, Hanken H. Does a preoperative cone beam CT reduce complication rates in the surgical removal of complex lower third molars? A retrospective study including 486 cases. *Head Face Med* 2021;17:33.
9. Hua J, Aziz S, Shum JW. Virtual surgical planning in oral and maxillofacial surgery. *Oral Maxillofac Surg Clin* 2019;31:519–30.
10. Zhang HX, Yan ZY, Cui NH, Sun F, Wu BZ. Accuracy of computer-assisted dynamic navigation when performing coronectomy of the mandibular third molar: a pilot study. *J Dent* 2023;139:104762.
11. Fangfang X, Yuxin G, Ahmadi S, et al. A prospective randomized study on the efficacy of real-time dynamic navigation in deep horizontal mandibular third molar extractions. *BMC Oral Health* 2024;24:1234.
12. Emery RW, Korj O, Agarwal R. A review of in-office dynamic image navigation for extraction of complex mandibular third molars. *J Oral Maxillofac Surg* 2017;75:1591–600.
13.. Feng J, Qi W, Duan S, et al. Three-dimensional printed model of impacted third molar for surgical extraction training. *J Dent Educ* 2021;85(12):1828–36.
14. Zeng J, Jiang Y, Liu Y, Hu Y, Guo C, Zhou L. Research on a novel digital tooth sectioning guide system for tooth sectioning during mandibular third molar extraction: an in vitro study. *J Stomatol Oral Maxillofac Surg* 2023;124:101383.
15. Szalma J, Lovász BV, Lempel E, Maróti P. Three-dimensionally printed individual drill sleeve for depth-controlled sections in third molar surgery. *J Oral Maxillofac Surg* 2019;77:704.e1–7.
16. Weissheimer A, Menezes LM, Sameshima GT, Enciso R, Pham J, Grauer D. Imaging software accuracy for 3-dimensional analysis of the upper airway. *Am J Orthod Dentofacial Orthop* 2012;142:801–13.
17. Lo Giudice A, Ronsivalle V, Grippaudo C, et al. One step before 3d printing-evaluation of imaging software accuracy for 3-dimensional analysis of the mandible: a comparative study using a surface-to-surface matching technique. *Materials* 2020;13:2798.
18. Didziokas M, Pauws E, Kölby L, Khonsari RH, Moazen M. BounTI (boundary-preserving threshold iteration): a user-friendly tool for automatic hard tissue segmentation. *J Anat* 2024;245:829–41.
19. Su S, Jia X, Zhan L, Gao S, Zhang Q, Huang X. Automatic tooth periodontal ligament segmentation of cone beam computed tomography based on instance segmentation network. *Heliyon* 2024;10:e24097.
20. Kot WY, Au Yeung SY, Leung YY, Leung PH, Yang WF. Evolution of deep learning tooth segmentation from CT/CBCT images: a systematic review and meta-analysis. *BMC Oral Health* 2025;25:800.
21. Huang C, Wang Y, Wang Y, Zhao Z. Automatic detection and proximity quantification of inferior alveolar nerve and mandibular third molar on cone-beam computed tomography. *Clin Oral Invest* 2024;28:648.
22. Liu MQ, Xu ZN, Mao WY, et al. Deep learning-based evaluation of the relationship between mandibular third molar and mandibular canal on CBCT. *Clin Oral Invest* 2022;26:981–91.
23. Chun SY, Kang YH, Yang S, et al. Automatic classification of 3D positional relationship between mandibular third molar and inferior alveolar canal using a distance-aware network. *BMC Oral Health* 2023;23:794.
24. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* 2012;30:1323–41.
25. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imag* 2015;15:29.
26. Pinter C, Lasso A, Wang A, Jaffray D, Fichtinger G. SlicerRT: radiation therapy research toolkit for 3D slicer. *Med Phys* 2012;39:6332–8.
27. Polizzi A, Quinzi V, Ronsivalle V, et al. Tooth automatic segmentation from CBCT images: a systematic review. *Clin Oral Invest* 2023;27:3363–78.
28. Butnaru OM, Tatarciuc M, Luchian I, et al. AI efficiency in dentistry: comparing artificial intelligence systems with human practitioners in assessing several periodontal parameters. *Medicina (Kaunas)* 2025;61:572.
29. Talmazov G, Bencharit S, Waldrop TC, Ammoun R. Accuracy of implant placement position using nondental open-source software: an in vitro study. *J Prosthodont* 2020;29:604–10.
30. Balki I, Amirabadi A, Levman J, et al. Sample-size determination methodologies for machine learning in medical imaging research: a systematic review. *Can Assoc Radiol J* 2019;70:344–53.
31. Schneider J, Vlachos M, eds. *Personalization of deep learning. Data science – analytics and applications*. Wiesbaden: Springer Fachmedien Wiesbaden, 2021.
32. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the aapm radiation therapy committee task group no. 132. *Med Phys* 2017;44:e43–76.
33. Deng HH, Liu Q, Chen A, et al. Clinical feasibility of deep learning-based automatic head CBCT image segmentation and landmark detection in computer-aided surgical simulation for orthognathic surgery. *Int J Oral Maxillofac Surg* 2023;52:793–800.
34. Irshad TB, Pascoletti G, Bianconi F, Zanetti EM. Mandibular bone segmentation from CT scans: quantitative and qualitative comparison among software. *Dent Mater* 2024;40:e11–22.
35. Bao J, Tan Z, Sun Y, et al. Deep ensemble learning-driven fully automated multi-structure segmentation for precision craniomaxillofacial surgery. *Front Bioeng Biotechnol* 2025;13:1580502.